

# Testes Diagnósticos

Prof. Dr. Maurício de Andrade Pérez<sup>1</sup>

**Sumário:** Embora médicos apresentem certa resistência quanto aos diversos usos da matemática, os exames diagnósticos fornecem um arcabouço extremamente rico para a interpretação de exames bem como o estabelecimento de critérios diagnósticos.

Neste texto, serão discutidas as probabilidades de acerto dos exames na população (sensibilidade e especificidade), dentre os indivíduos examinados (valores preditivo positivo e negativo), a avaliação dos próprios exames (razão de verossimilhança positiva e negativa), e a aplicação destes em situações de emergência (testes em paralelo) e no âmbito ambulatorial (testes em série).

Espera-se que com esta leitura, o aluno e/ou médico possa visualizar de forma crítica os exames solicitados, bem como a maneira que diversas rotinas de diagnóstico são estabelecidas.

**Abstract:** *Albeit physicians may withstand mathematical issues, probabilities applied to diagnostic testing are a major topic in contemporary medicine. This article will argue on the connections regarding diagnostic testing and epidemiology, using the many probabilities that may be calculated to assert an exam on a given population (sensitivity and specificity), issued by a positive or negative exam (predicted value positive or negative), to appraise the exam itself (positive or negative likelihood ratio) as well as to evaluate diagnostic testing in emergencies or outpatient settings.*

*We hope that among these pages, the medical student or physician may strengthen his grasp on how and why a given diagnostic routine is applied.*

Uma das muitas interfaces entre a epidemiologia e a clínica reside na análise crítica do processo de exames e diagnósticos. Este capítulo parte da premissa que o pedido de exames bem como o raciocínio clínico são basicamente processos probabilísticos de acertos e erros, podendo assim ser equacionado. Imagine um paciente entrando numa emergência com dor retroesternal com irradiação para o antebraço, dispnéia, ritmo irregular etc. Este quadro representaria de forma muito mais provável uma síndrome coronariana do que qualquer outra nosologia (probabilidade esta que é intuitiva, mas não obstante uma probabilidade).

Iniciaremos nossa discussão examinando a tabela 1.1 e a figura 1.1 abaixo.

TABELA 1.1

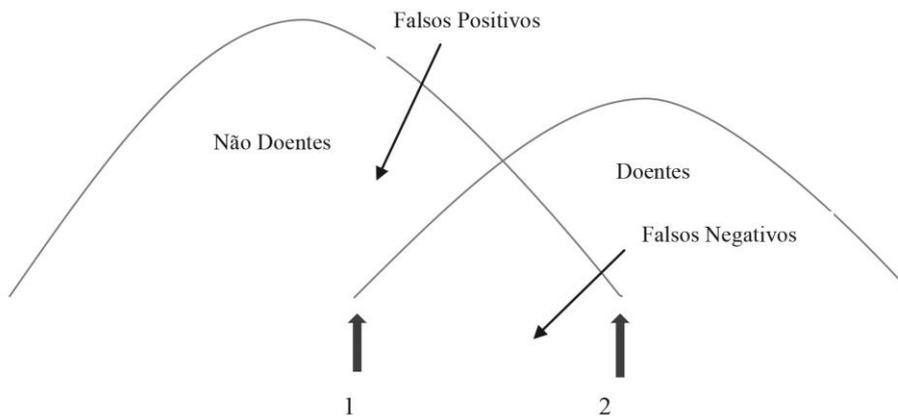
**Comparação entre os resultados de um diagnóstico de certeza (padrãoouro), com um exame utilizado para avaliar uma dada população (N)**

	PADRÃO-OURO (EXAME DE CERTEZA)		
Exame	Doente	Não Doente	Total
Doente	a	b	a+b
Não Doente	c	d	c+d
Total	a+c	b+d	N

FIGURA 1.1

**Representação das relações entre uma população doente e não doente**

<sup>1</sup> Professor Assistente da Escola de Medicina Souza Marques e da Universidade Federal do Rio de Janeiro



Pela tabela 1.1 e figura 1.1 podemos deduzir quatro diferentes probabilidades, sendo que neste momento iremos nos concentrar em duas, a sensibilidade e a especificidade de um exame diagnóstico.

Inicialmente vamos definir o que seria um padrão-ouro, também conhecido como exame (ou exames) de certeza, ou ainda como a “verdade”. A lógica básica reside no fato de que ao criarmos um exame qualquer (nova prova de função hepática, novo diagnóstico de imagem, novo aparelho para verificar a pressão arterial etc.), este exame deverá ser comparado com a “verdade” (padrão-ouro).

Quando eu estava na graduação, participei de um projeto para criar uma definição clínica para AIDS, isto é, a partir de determinados sinais e sintomas, o paciente seria classificado ou não como possuindo a doença. O padrão-ouro (“verdade”) foi estabelecido como todo paciente que possuísse três exames de Elisa, Imunofluorescência, Western Blot e cultura de vírus positiva. O “não doente” seria aquele com todos estes exames negativos. Qualquer indivíduo que não cumprisse estas condições seria excluído do estudo (que ficou conhecido como definição Rio de Janeiro/Caracas).

A sensibilidade é definida como a probabilidade de acerto de um exame dentre as pessoas que possuem a doença, enquanto a especificidade a probabilidade de acerto do exame entre os que não a possuem.

O exame da tabela 1.1 nos fornece portanto duas fórmulas iniciais, onde a sensibilidade será calculada como  $\frac{a}{a+c}$  e a especificidade como  $\frac{d}{b+d}$ .

Ainda pela tabela 1.1, podemos perceber a existência de dois erros fundamentais. Em “b” uma pessoa normal foi classificada como sendo doente (falso positivo) e em “c” uma pessoa doente foi classificada como normal (falso negativo). Uma questão que surge, portanto seria qual dos dois erros seria mais importante. Esta não é uma pergunta trivial e nem simples de responder, pois para uma dada doença, os falsos positivos deveriam ser evitados a qualquer custo, e numa outra situação (para a mesma doença), a opção seria evitar os falsos negativos. Imagine por exemplo que uma pessoa não doente fosse diagnosticada como sendo HIV+ ou ainda com uma massa tumoral no mediastino. Em ambos os casos o paciente sofreria uma completa revolução em sua vida, sendo que no segundo caso teríamos o mesmo sendo exposto a uma cirurgia (com todos os riscos decorrentes desta), sem que uma doença de fato existisse. Em outras palavras temos que este erro simplesmente não pode ocorrer.

Alternativamente, em “c” temos um paciente HIV+ doando sangue onde sua doença não foi detectada pelo banco de sangue, ou ainda um indivíduo com uma massa tumoral não identificada, perdendo-se assim uma janela de oportunidade extremamente pequena que poderia alterar completamente o prognóstico do mesmo.

Complementando esta discussão, examinemos a figura 1.1 acima exposta. Para melhor compreender sua leitura, imaginemos uma definição diagnóstica para a hipertensão arterial. Digamos que esta definição assuma que o paciente será hipertenso se sua pressão arterial, verificada três vezes por dia, num período de 30 dias (com um total de 90 medidas), estivesse sempre abaixo de 120 (sistólica) por 80 (diastólica). Esta situação refletiria o ponto de corte “1” (figura 1.1) que significaria na prática a capacidade de diagnosticar 100% dos doentes. De fato, é virtualmente impossível que um paciente hipertenso, não tratado, apresente uma pressão normal durante um mês inteiro de avaliação (reparar que pela definição acima, uma medida de 121 x 70, seria considerada “anormal”). Como todos os hipertensos foram diagnosticados, o erro “c” (falso negativo) não existiria, o que faria com que a sensibilidade fosse calculada como 100%. Em outras palavras temos que um exame 100% sensível seria aquele capaz de diagnosticar todos (100%) dos doentes.

Pela mesma figura, observe que o ponto de corte “1”, faria também com que o número de falsos positivos (pessoas não hipertensas sendo diagnosticadas como doentes) seria imenso, diminuindo em muito a especificidade.  $\frac{a}{a+c} =$

Alternativamente, criemos agora outra definição diagnóstica, onde um indivíduo seria classificado como hipertenso apenas se a pressão arterial (verificada três vezes por dia, durante um mês), estivesse sempre acima de 210 (sistólica) e 130 (diastólica). Temos agora, uma situação oposta à anterior, onde seria virtualmente impossível que um paciente normotenso, apresentasse continuamente ao longo de um mês estes níveis pressóricos. O erro “b” (falsos positivos) seria igual a zero, o que faria com que a especificidade fosse calculada como  $\frac{a}{a} = 100\%$ . Desta forma o teste não acusaria como doente nenhum paciente normotenso. Por outro lado, a quantidade de indivíduos hipertensos (leves e moderados) que seriam também classificados como “normais”, estaria aumentando em muito o erro nos falsos negativos, baixando enormemente a sensibilidade.

Desta discussão podemos concluir que ao buscarmos uma maior sensibilidade, perderemos em especificidade e vice-versa. Para avançar um pouco mais nesta discussão iremos agora calcular para um determinado exame (prova de esforço) as consequências de sua aplicação em diferentes populações (tabelas 1.2-1.4).

Iniciaremos esta discussão pela análise da tabela 1.2 abaixo. Ela representa os resultados de uma meta-análise (Gianrossi, et al., 1989) (Kwok, et al., 1999) englobando centenas de diferentes artigos que foram agrupados segundo as características dos doentes. A importância desta tabela é a de demonstrar a utilidade de uma meta-análise, onde os artigos seriam avaliados por uma equipe de especialistas (epidemiologistas, matemáticos, clínicos etc.) facilitando assim o trabalho do médico. Basicamente temos que a sensibilidade (nas diversas situações) estaria variando entre 68-72%, enquanto que a especificidade entre 69-84% (excluí um grupo de três artigos que apontavam para uma especificidade de 50 %, exatamente por serem muito diferentes dos demais, além de serem estudos que continham pouquíssimos pacientes). No exemplo a seguir iremos definir a sensibilidade como 70% e a especificidade como 80%.

TABELA 1.2  
Meta-análise da prova de esforço

Classificação	Nº de estudos	Total de pacientes	Sensibilidade	Especificidade
Meta-análise da Prova de esforço (padrão)	147	24.047	68 %	77 %
Meta-análise sem IAM	58	11.691	67 %	72 %

Meta-análise (estudos sem viés)	3	> 1.000	50 %	50 %
Meta-análise com desnível ST	22	9.153	69 %	70 %
Meta-análise sem desnível ST	3	840	67 %	84 %
Meta-análise com digoxina	15	6.338	68 %	74 %
Meta-análise sem digoxina	9	3.548	72 %	69 %
Meta-análise com hipertrofia ventricular esquerda	15	8.016	68 %	69 %
Hipertrofia sem hipertrofia ventricular esquerda	10	1.977	72 %	77 %

As tabelas 1.3<sub>a-d</sub> e 1.4<sub>a-d</sub> serão construídas baseando-se em duas situações clínicas distintas, no que refere à prova de esforço. No primeiro grupo (1.3<sub>a-d</sub>) o exame será aplicado numa população com média de idade de 50 anos, incluindo predominantemente indivíduos hipertensos, diabéticos, com colesterol elevado à custa do LDL, sedentários e fumantes e que encontram numa consulta ao cardiologista, pois estão interessados em ingressar numa academia de ginástica. Nas tabelas 1.4<sub>a-d</sub>, teremos uma população jovem, com média de idade ao redor de 18 anos, sem qualquer fator de risco ou dado que indicasse uma possível doença coronariana. Podemos inferir, portanto que o conjunto de tabelas 1.3<sub>a-d</sub> estaria analisando uma população com alta prevalência de doença (também conhecida como probabilidade pré-teste) e que iremos situar ao redor de 30%. No segundo conjunto de tabelas teríamos ao contrário, uma baixíssima probabilidade pré-teste. Para facilitar nossa discussão, a prevalência neste segundo caso será estabelecida como sendo 1% (o que é um exagero, pois estaríamos dizendo que um jovem a cada 100 apresentaria uma doença coronariana) apenas para facilitar nossa discussão. Finalmente, teremos que o teste será realizado em 100.000 pessoas para cada conjunto de tabelas. Temos então:

TABELA 1.3a  
Montagem inicial da tabela (alta prevalência).

	Padrão-ouro (exame de certeza)		
Exame	Doente	Não Doente	Total
Doente	a	b	a+b
Não Doente	c	d	c+d
Total	40.000 (a+c)	60.000 (b+d)	100.000 (N)

Se a prevalência é de 40 % (ou seja, 0,40) e a população total é de 100.000 indivíduos, teremos então que o número total de doentes (a + c) será de  $0,40 \times 100.000 = 40.000$ . Se o total de indivíduos é de 100.000 e existem 40.000 doentes, temos então que o total de sadios (b + d) será de  $100.000 - 40.000$ , ou seja, 60.000 pessoas. Examinemos agora a sensibilidade (70 %) e a especificidade (80 %) deste exame.

TABELA 1.3b  
Montagem inicial da tabela (alta prevalência), sensibilidade.

	Padrão-ouro (exame de certeza)		
Exame	Doente	Não Doente	Total

<b>Doente</b>	28.000 (a)	b	a+b
<b>Não Doente</b>	12.000 (c)	d	c+d
<b>Total</b>	40.000 (a+c)	60.000 (b+d)	100.000 (N)

Sabemos que a sensibilidade é de 70 % e que esta é calculada como  $\frac{a}{a+c}$  teremos, portanto que  $0,70 = \frac{a}{40.000} \Rightarrow a = 40.000 \times 0,70 = 28.000$  e que “c” será  $40.000 - 28.000 = 12.000$ . Perceba que este exame, com esta sensibilidade, permitiu que existissem 12.000 falsos negativos (doentes não detectados pelo teste de esforço). Continuando a construção da tabela, iremos inserir agora a especificidade (tabela 1.3c).

TABELA 1.3c

**Montagem inicial da tabela (alta prevalência), especificidade.**

	<b>Padrão-ouro (exame de certeza)</b>		
<b>Exame</b>	Doente	Não Doente	Total
<b>Doente</b>	28.000 (a)	6.000 b	a+b
<b>Não Doente</b>	12.000 (c)	54.000 d	c+d
<b>Total</b>	40.000 (a+c)	60.000 (b+d)	100.000 (N)

Como a especificidade é de 80% e seu cálculo é feito pela fórmula  $\frac{d}{b+d}$  teremos que  $0,80 = \frac{d}{60.000} \Rightarrow d = 0,80 \times 60.000 = 48.000$  e que  $b =$

$60.000 - 48.000 = 12.000$  (falsos positivos). Podemos agora refletir sobre esses dois erros. O falso negativo (12.000 doentes) é certamente indesejável. Para diminuí-lo deveremos buscar um exame mais sensível. Todavia o aumento da sensibilidade (ver figura 1.1) irá aumentar o número de falsos positivos. Neste momento é fundamental que se raciocine qual dos erros é o mais importante, em outras palavras, qual a consequência de cada um. Provavelmente o exame a seguir após uma prova de esforço positiva (alterada) será uma cinecoronariangiografia, que, há 20 anos, apresentava uma mortalidade de aproximadamente 1%. Apesar de não ser a situação ideal, perder 1% dos 28.000 doentes, com uma doença grave, que necessitam de uma intervenção razoavelmente urgente, seria um preço razoável a pagar. A questão é que se esta mortalidade for semelhante nos falsos positivos (6.000 indivíduos saudáveis) temos agora um evento absolutamente indesejável. Diminuir os falsos positivos, entretanto (aumentar a especificidade) implicaria numa diminuição da sensibilidade (perder mais do que 12.000 doentes). Há 20 anos, a decisão foi tomada pelo equilíbrio dos dois erros, investindo-se em rotinas que aumentassem a segurança da cinecoronariangiografia. Temos hoje, uma quantidade muito maior de falsos positivos do que no passado, uma vez que o exame se tornou mais seguro.

Para fechar a tabela, calcularemos agora a última parte. Existirão, portanto  $28.000 + 6.000 = 34.000$  indivíduos tidos como positivos pela prova de esforço e  $12.000 + 54.000 = 66.000$  negativos por este exame. Como mencionado acima, a “verdade” será dada por um exame de certeza, realizado por artigo especialmente direcionado à análise do mesmo. Por outro lado, o médico não tem acesso a esta “verdade”.

TABELA 1.3d

**Montagem inicial da tabela (alta prevalência), totalização do exame.**

Exame	Padrão-ouro (exame de certeza)		Total
	Doente	Não Doente	
<b>Doente</b>	28.000 (a)	6.000 b	34.000 a+b
<b>Não Doente</b>	12.000 (c)	54.000 d	66.000 c+d
<b>Total</b>	40.000 (a+c)	60.000 (b+d)	100.000 (N)

O que o clínico teria de fato é que a cada 34.000 exames positivos, haveria acerto em 28.000 deles, ou seja,  $\frac{a}{a+b} = \frac{28.000}{34.000} = 82,4\%$ , valor este que é denominado de Valor Preditivo Positivo (VPP ou VP+). Em outras palavras, deveríamos “arriscar” uma conduta (exame) mais agressiva, pois a probabilidade das pessoas detectadas pela prova de esforço, realmente possuírem uma síndrome coronariana é alta (82,4%).

Temos ainda que das 66.000 pessoas consideradas como negativas,

$\frac{54.000}{66.000} = 81,8\%$  seria a probabilidade de acertadamente considerá-las como negativas. Como se pode verificar, esta conduta conseguiu um equilíbrio entre os dois erros (falsos positivos e falsos negativos).

Analisemos agora o mesmo exame numa prevalência muito mais baixa (1 %). A lógica deste raciocínio é a de que a sensibilidade e especificidade são propriedades razoavelmente fixas dos exames, isto é, caso ele seja aplicado de forma tecnicamente correta, não deveremos observar alterações significativas em relação à sua sensibilidade e especificidade. Analogamente ao raciocínio anterior teremos que se a prevalência é de 1 % (0,01) e que o total de indivíduos testados é de 100.000, o cálculo da parte inferior da tabela será dado por  $a + c = 0,01 \times 100.000 = 1.000$  (total de doentes). Concluindo esta parte, temos que  $b + d$  (total de não doentes) =  $100.000 - 1.000 = 99.000$ .

Como dito anteriormente a sensibilidade e especificidade deverão manter suas margens de acerto e de erro. Desta forma:  $0,70$  (Sensibilidade) =  $\frac{a}{1.000} \Rightarrow a = 0,70 \times 1.000 = 700$  e  $c = 1.000 - 700 = 300$

TABELA 1.4a

**Montagem inicial da tabela (baixa prevalência).**

Exame	Padrão-ouro (exame de certeza)		Total
	Doente	Não Doente	
<b>Doente</b>	a	b	a+b
<b>Não Doente</b>	c	d	c+d
<b>Total</b>	1.000 (a+c)	99.000 (b+d)	100.000 (N)

TABELA 1.4b

**Montagem inicial da tabela (baixa prevalência), sensibilidade.**

Exame	Padrão-ouro (exame de certeza)		Exame
	Doente	Não Doente	
<b>Doente</b>	700 a	b	Doente
<b>Não Doente</b>	300 c	d	Não Doente

<b>Total</b>	1.000 (a+c)	99.000 (b+d)	100.000 Total
--------------	-------------	--------------	---------------

TABELA 1.4c

**Montagem inicial da tabela (baixa prevalência), especificidade.**

	Padrão-ouro (exame de certeza)		
<b>Exame</b>	Doente	Não Doente	Exame
<b>Doente</b>	700 a	19.800 b	Doente
<b>Não Doente</b>	300 c	79.200 d	Não Doente
<b>Total</b>	1.000 (a+c)	99.000 (b+d)	100.000 Total

Continuando a construção da tabela teremos:

Especificidade =  $\frac{d}{b+d}$  0,80 =  $\frac{d}{99.000}$   $\Rightarrow d = 0,80 \times 99.000 = 79.200$  e que  $b = 99.000 - 79.200 = 19.800$ .

Surpreendentemente, apesar de termos utilizado **exatamente** a mesma sensibilidade (70 %) e especificidade (80 %) nas tabelas onde a prevalência era mais alta, o número erros representados pelos falsos negativos (300 agora e 12.000 anteriormente) e falsos positivos (19.800 contra 6.000 na alta prevalência) são completamente diferentes. A razão para este fato reside exatamente na prevalência da doença em questão. Apesar da sensibilidade não ter sido modificada, repare que se existe uma probabilidade de 70 % de acerto (sensibilidade) então teremos 30 % de erro. Esta chance de erro numa população muito grande (pois a prevalência elevada significa que teremos muitos doentes) representará uma quantidade enorme de falsos negativos na primeira série de tabelas (1.3<sub>a-d</sub>). Por outro lado, a mesma sensibilidade (70 % de acerto e 30 % de erro) acarretará em muito menos falsos negativos quando a prevalência for baixa. Em outras palavras temos o mesmo erro de 30 %. Entretanto 30 % de muitos pacientes (prevalência alta) terá como consequência muitos falsos negativos e este mesmo erro numa prevalência baixa representa um número absoluto muito menor, apesar de ser exatamente a mesma sensibilidade. Raciocínio análogo poderá ser feito para os falsos positivos. Na primeira série de tabelas (1.3<sub>a-d</sub>) temos menos pessoas “não doentes” (pois se a prevalência é alta temos menos pessoas “normais”) do que na segunda série (1.4<sub>a-d</sub>) onde por definição uma baixa prevalência significa muitos “não doentes”. Se a chance de acerto (especificidade neste caso) é de 80 %, teremos 20 % de erro (no grupo de indivíduos sem doença). Esta probabilidade (20 %) aplicada a uma enorme população sem doença (prevalência baixa) irá acarretar uma quantidade bem maior de falsos positivos, apesar da especificidade ser exatamente a mesma em ambas as situações. O que isto significa então? Na tabela 1.4<sub>d</sub> podemos agora fechar nosso raciocínio, bem como as implicações clínicas deste exame numa população como baixa prevalência.

TABELA 1.4d

**Montagem inicial da tabela (baixa prevalência), totalização do exame**

	Padrão-ouro (exame de certeza)		
<b>Exame</b>	Doente	Não Doente	Exame
<b>Doente</b>	700 a	19.800 b	20.500 (a + b)
<b>Não Doente</b>	300 c	79.200 d	79.500 (c + d)

<b>Total</b>	1.000 (a+c)	99.000 (b+d)	100.000 Total
--------------	----------------	--------------	---------------

O total de positivos pelo exame será de  $700 + 19.800 = 20.500$  e o total de negativos é dado por  $300 + 79.200 = 79.500$ . O Valor preditivo positivo (VPP) pode ser calculado como:

$$\text{VPP} = \frac{700}{20.500} = 0,034 \text{ ou } 3,4\%$$
Em outras palavras, a chance de um exame positivo neste caso estar correto é de apenas 3,4% (lançar uma moeda para cima onde “cara” diagnosticaríamos a doença e “coroa” o não doente, possui uma chance de acerto de 50%) o que implicaria que na prática o exame não serviu para orientar o médico, diante um resultado positivo.

Alternativamente, o Valor preditivo negativo (VPP) será 
$$\text{VPP} = \frac{79.200}{79.500} = 99,6\%$$
Temos de forma complementar que um resultado negativo possui 99,6% de chance de acerto. Conclusão: Diante de um resultado positivo o médico não deverá confiar no mesmo e diante de um teste negativo ele saberá algo que já conhecia previamente (praticamente ninguém nesta população possui doença coronariana). A partir deste momento seria lícito questionar qual o porquê deste exame ter sido solicitado.

É importante também perceber que esta discussão também toca a própria alma de uma boa anamnese e exame físico. Quando o médico coleta a história da doença atual e adiciona a esta o exame físico ele de fato estará iniciando a construção de uma hipótese diagnóstica baseada em sinais e sintomas (aumento da prevalência de determinada nosologia), onde um resultado de exame positivo terá valor completamente diferente de quando o exame é feito sem qualquer critério.

Temos que até este momento os principais pontos discutidos foram:

- A sensibilidade é a probabilidade de corretamente classificar um indivíduo como doente, na população verdadeiramente doente.
- A especificidade é a probabilidade de corretamente classificar um indivíduo como não doente, na população sem a doença.
- O valor preditivo positivo representa a probabilidade de acerto do teste diagnóstico, dentre aqueles classificados como doentes (pelo teste).
- O valor preditivo negativo representa a probabilidade de acerto do teste diagnóstico, dentre aqueles classificados como não doentes (pelo teste).
- A baixa prevalência da doença numa população tende a aumentar o número de falsos positivos.
- A elevada prevalência da doença numa população tende a aumentar o número de falsos negativos.

Vamos examinar agora uma forma de avaliar o próprio teste, determinando ou não o seu valor, bem como estabelecendo um método que me permita comparar dois ou mais testes diagnósticos para verificar qual seria mais interessante.

### Razão de verossimilhança

Existem duas maneiras básicas de se avaliar um teste: através da razão de verossimilhança positiva e negativa ou ainda através de um método gráfico (curva ROC - “Receiver Operating Curve”).

A razão de verossimilhança positiva (RV+) é dada por 
$$\frac{\text{Sensibilidade}}{1 - \text{Especificidade}}$$
e a razão de verossimilhança negativa (RV-) por 
$$\frac{1 - \text{Sensibilidade}}{\text{Especificidade}}$$
Vamos raciocinar um pouco acerca das duas fórmulas apresentadas. Na RV+ temos no numerador a sensibilidade (probabilidade de acerto do teste entre os verdadeiros doentes). Por outro lado o denominador não é a especificidade e sim “1

- especificidade”. Ora, a especificidade é a chance de acerto do teste na população sem doença (verdadeiramente). Então “1 - especificidade” é o complemento desta, ou seja, a chance de erro entre os verdadeiramente não doentes. Se a especificidade é de 0,8 (chance de acerto) então isto implica que  $1 - 0,8 = 0,2$  é a minha chance de erro (na população sem doença). Na prática termos, portanto que a  $RV+$  está dividindo uma probabilidade de acerto entre os doentes (sensibilidade) por uma probabilidade de erro entre os não doentes (1 - especificidade). Podemos concluir que quanto menos erros houver (ou mais acertos), maior será a  $RV+$ . Exemplo:  $RV+ = 5 \Rightarrow$  O exame “acerta” em cinco indivíduos classificados como tendo a doença para cada erro;  $RV+ = 20 \Rightarrow$  O exame “acerta” em 20 indivíduos classificados como tendo a doença para cada erro e assim por diante.

Examinemos agora a razão de verossimilhança negativa ( $RV-$ ). Sua fórmula compreende a especificidade no denominador (chance de acerto do exame nos indivíduos verdadeiramente sadios). Todavia o numerador é dado por “1 - sensibilidade”. Analogamente ao exemplo anterior, se a sensibilidade é a probabilidade de acerto do exame entre os verdadeiramente doentes, “1 - sensibilidade”, será seu complemento, ou seja, a chance de erro entre os doentes verdadeiros (ou ainda a proporção de falsos negativos em relação ao total de doentes). Temos, portanto que a  $RV-$  está sendo calculada como uma probabilidade de erro (1 - sensibilidade) no numerador e uma probabilidade de acerto (especificidade) no denominador. Quanto menos erros ocorrerem no numerador (1 - sensibilidade) menor será o resultado da  $RV-$ , ou seja, melhor o exame. Exemplos:  $RV- = 0,1 \Rightarrow$  a cada cem exames negativos, o teste erra em 10;  $RV- = 0,3 \Rightarrow$  a cada cem exames negativos o teste erra em 30, e assim por diante.

Vamos imaginar um teste (teste 1) com sensibilidade de 70 % e especificidade de 80 % e que custe cerca de R\$ 20,00. Digamos ainda que exista um teste 2 com sensibilidade de 85 % e especificidade de 90 %, ao custo de R\$ 100,00. Para facilitar nossa discussão, vamos assumir que ambos apresentem boa aceitação (adesão) dos pacientes e a mesma proporção de iatrogenias/efeitos colaterais. Obviamente um exame é mais sensível e específico do que o outro, mas você deseja convencer o diretor do hospital ou secretário de saúde que valeria a pena pagar um preço cinco vezes maior.

Exame 1:  $RV+ = \frac{0,70}{1-0,8} = \frac{0,70}{0,2} = 3,5 \Rightarrow$  dentre os pacientes classificados como doentes, o exame acerta em 3,5 pacientes para cada erro, ou ainda em 35 pacientes para cada 10 erros.

$RV- = \frac{1-0,70}{0,8} = \frac{0,30}{0,80} = 0,38 \Rightarrow$  a cada 100 exames negativos, este teste erra em 38.

Exame 2:  $RV+ = \frac{0,85}{1-0,9} = \frac{0,85}{0,1} = 8,5 \Rightarrow$  dentre os indivíduos classificados como doentes, o exame acerta em 8,5 pacientes para cada erro (ou 85 pacientes para cada 10 erros).

$RV- = \frac{1-0,85}{0,9} = \frac{0,15}{0,90} = 0,17 \Rightarrow$  a cada 100 exames negativos, este teste erra em 17.

Temos agora que o teste 2 pode ser expresso de forma palpável em termos de acertos e erros. O segundo exame é de fato mais caro, mas sua capacidade de acerto é muito maior. Juntando-se a importância do diagnóstico não demorar a ser firmado (maior letalidade? maior tempo de permanência/custo da internação?), podemos decidir baseados em um conjunto de argumentos a favor ou contra a compra de um ou outro método diagnóstico.

Como dito anteriormente, uma segunda forma de avaliar o próprio exame seria a sua representação gráfica (“Receiver Operating Curve” ou curva ROC). Esta forma de avaliação é facilmente encontrada em artigos que estejam avaliando um método de diagnóstico qualquer, sendo geralmente citados os dois métodos (razão de verossimilhança e a curva ROC).

A tabela 1.5 abaixo representa os diversos pontos de corte da glicemia pós-prandial numa população qualquer, bem como a sensibilidade e especificidade correspondentes. Os chamados “pontos de corte” de um exame podem facilmente ser determinados classificando uma população como doente/não doente para cada critério estabelecido. Basicamente o pesquisador estará avaliando a sensibilidade e especificidade resultante caso a glicemia normal seja estabelecida em diversos pontos,  $\leq 120$  ou  $\leq 130$  ou  $\leq 140$  etc.

TABELA 1.5  
**Representação da sensibilidade e especificidade da glicemia pós-prandial (Medronho, et al., 2002)**

Glicemia	Sensibilidade (%)	Especificidade (%)
70	98,6	8,8
80	97,1	25,5
90	94,3	47,6
100	88,6	69,8
110	85,7	84,1
120	71,4	92,5
130	64,3	96,9
140	57,1	99,4
150	50,0	99,6
160	47,1	99,8
170	42,9	100,0
180	38,6	100,0
190	34,3	100,0
200	27,1	100,0

A estimativa do melhor ponto de corte, que foi se modificando ao longo do tempo, é um fascinante exemplo de como a compreensão dos mecanismos envolvidos na sensibilidade e especificidade podem e devem ser utilizados no benefício dos pacientes.

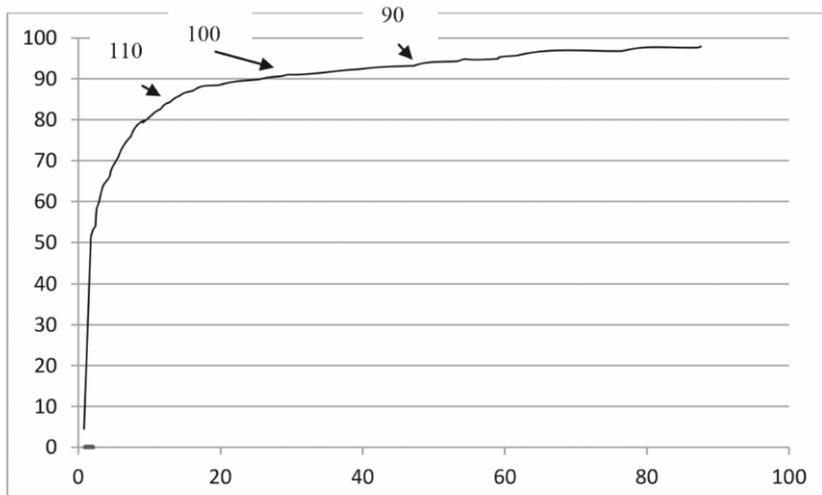
Há 60 anos, um ponto de corte considerado razoável seria o de 170 ou mais. Nesta situação temos uma especificidade de 100% (nenhuma pessoa sem diabetes teria este diagnóstico). Se levarmos em consideração que naquela época não havia adoçantes, gelatina dietética etc., faria todo sentido não cometer um erro falso positivo (o sacrifício de uma pessoa não doente seria imenso). Em contrapartida teríamos apenas 42,9 % de sensibilidade (o diagnóstico de um diabético estaria ocorrendo de forma atrasada em ao menos metade deles).

Há cerca de 30 anos, o ponto de corte migrou para 110 (ver curva ROC no gráfico 1 abaixo). De fato, em tese, este seria o melhor índice de normalidade, pois é o momento onde a sensibilidade e a especificidade teriam o máximo valor simultaneamente. Se baixarmos este ponto de corte para 100 ou 90, ganharemos em sensibilidade, mas a queda da especificidade seria imensa (respectivamente de 69,8 % e 47,6 %.). Atualmente considera-se como “intolerante à glicose” um paciente com glicemia de 100. A opção por este ponto de corte é claramente o de aumentar a sensibilidade em detrimento da especificidade. Particularmente, acredito que ele deveria ser deslocado para 90, onde a sensibilidade alcançaria 94%, mesmo com uma imensa perda na especificidade (47,6 %).

O processo decisório está passando pela seguinte questão: É fundamental um diagnóstico precoce (alta sensibilidade), por outro lado, a abordagem clínica deste paciente (com “intolerância

à glicose”) seria uma dieta para o excesso de sobrepeso (que seria importante para qualquer indivíduo do ponto de vista cardiovascular, ósteo-articular, etc.), sendo o mesmo diabético ou não. Além disto, o médico incentivaria seu paciente a realizar atividades físicas ao menos três vezes por semana (novamente sem malefícios), ao mesmo tempo que a oferta de produtos dietéticos é imensa nos dias atuais.

GRÁFICO 1.1  
Curva ROC da glicemia pós-prandial (Medronho, et al., 2002)

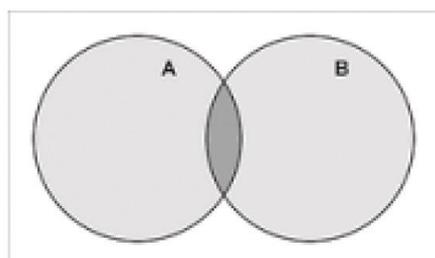


### Testes em Série e Testes em paralelo

Como dito anteriormente, o raciocínio médico é basicamente probabilístico. Imagine um paciente chegando numa emergência com dor retroesternal com irradiação para o braço, sudorese, ritmo irregular, hipotensão, dispneia e um eletrocardiograma (ECG) normal. Nesta situação, apesar do ECG normal, o médico dificilmente deixará de pensar num enfarto. Imaginemos ainda uma criança com febre, desorientação, vômitos em jato, rigidez de nuca com punção de liquor de aspecto normal. Apesar do aspecto do liquor, dificilmente descartaríamos de forma imediata um diagnóstico de meningite.

Em ambos os casos, estamos diante de uma situação que a epidemiologia denomina de um teste em paralelo, ou seja, numa emergência o maior erro que se pode cometer é o de não tratar e/ou acompanhar um paciente com uma patologia com real risco de vida (falso negativo), o que em outras palavras quer basicamente dizer que a sensibilidade do diagnóstico em muito suplanta a especificidade. Se considerarmos que a sensibilidade de cada exame representa um conjunto, temos então pela figura 2 que:

FIGURA 1.2  
União de dois conjuntos



A sensibilidade dos dois exames pode ser considerada como a união de dois conjuntos, cada um representando um dos exames propriamente ditos. Matematicamente podemos inferir que  $A \cup B$  seria dada pela soma do acerto do primeiro exame com a do segundo, menos a interseção (que está sendo contada duas vezes, uma quando o exame A foi positivo que se junta na mesma área em que o exame B também é positivo - ver a parte mais escura da figura 2). Em termos matemáticos temos:  $A \cup B = \text{Sensibilidade (exame A)} + \text{Sensibilidade (exame B)} - \text{a interseção dos dois}$ , ou seja,  $A \cup B = \text{Sens}_A + \text{Sens}_B - \text{Sens}_A \times \text{Sens}_B$ .

Por outro lado, para que um paciente seja liberado da emergência, todos os exames deverão apresentar valores dentro da normalidade, ou seja, no nosso exemplo do infarto, a clínica e o ECG não apresentariam alteração. Se observarmos novamente a figura 2, estamos exatamente na área hachurada, o que significa que a liberação do paciente será dada com todos os exames negativos. A especificidade (a decisão de que o paciente não apresenta a doença em questão) será dada por  $A \cap B$ , ou ainda Especificidade do primeiro exame x Especificidade do segundo exame, ou ainda  $\text{Esp}_A \times \text{Esp}_B$ .

Suponha que um primeiro exame possua sensibilidade de 60,0 % e especificidade de 70,0%, enquanto um segundo procedimento possua sensibilidade e especificidade de 70,0 % e 80,0 % respectivamente. Em paralelo (numa situação de emergência) em que qualquer resultado positivo seria aceito, temos:

$0,60 + 0,70 - 0,60 \times 0,70 = 1,30 - 0,42 = 0,88$  ou 88,0 %. Simultaneamente a especificidade seria então de:  $0,70 \times 0,80 = 0,56$  ou 56,0 %. Perceber, portanto que sacrificamos a especificidade (está muito mais baixa que qualquer dos exames separadamente), para favorecer a sensibilidade.

O interessante é que esta lógica pode ser usada em distintas situações. Digamos que a eficácia de um anticoncepcional hormonal seja de 96 % enquanto que um segundo método (digamos uma barreira física como o preservativo) seja de 97 %. O ginecologista ao prescrever os dois métodos estaria “acertando” duplamente. Por um lado, existiria a prevenção das doenças sexualmente transmissíveis. Mas ao utilizar dois métodos o médico estará lançando mão de uma técnica em paralelo, isto é, basta que um dos anticoncepcionais funcione (ele estará somando estas probabilidades), tendo então: Probabilidade de funcionamento (Hormônio) + Probabilidade de funcionamento (Preservativo) - quando os dois funcionam ao mesmo tempo. Matematicamente podemos calcular:  $0,96 + 0,97 - 0,96 \times 0,97 = 1,93 - 0,93 = \underline{\underline{100,0 \% !!!}}$

No ambulatório por sua vez (testes em série) a lógica clínica é bastante distinta. De fato, para que um paciente seja submetido a uma terapia mais agressiva (cirurgias complexas, exames muito invasivos, etc.), o médico gostaria de ter certeza da necessidade dos referidos procedimentos. Como estamos num ambulatório e não numa emergência, há tempo suficiente para se pedir diversos exames até que estejamos razoavelmente seguros da necessidade de uma maior intervenção.

Voltando à figura 2, a certeza estará ocorrendo quando os exames acusarem o mesmo diagnóstico, ou seja, estamos na interseção dos exames (a parte hachurada da figura), que nos dará a sensibilidade. Alternativamente, enquanto um ou mais exames forem negativos, continuamos a investigação, pois num ambulatório o falso positivo (especificidade) é mais importante, já que não desejamos que um indivíduo sem a doença (falso positivo) seja exposto a um procedimento de alto risco.

Analogamente ao raciocínio anterior temos que a sensibilidade será dada por: Sensibilidade (exame 1) x Sensibilidade (exame 2) ou ainda  $\text{Sens}_1 \times \text{Sens}_2$  (a interseção dos conjuntos). A especificidade por outro lado será calculada como: Especificidade (exame 1) + Especificidade (exame 2) - Especificidade (exame 1) x Especificidade (exame 2), o que representaria a união dos conjuntos.

Repetindo o mesmo exemplo dos testes em paralelo, onde o primeiro exame possuía sensibilidade de 60,0 % e especificidade de 70,0 % enquanto o segundo exame apresentava-se com sensibilidade e especificidade respectivamente de 70,0 % e 80,0 %, a opção por exames em série acarretaria em  $0,60 \times 0,70 = 42,0$  % de sensibilidade (muito mais baixa que qualquer dos exames separadamente) e  $0,70 + 0,80 - 0,70 \times 0,80 = 1,50 - 0,56 = 94,0$  % de especificidade.

Podemos finalizar concluindo que toda esta discussão nos permite justificar o porquê de uma determinada rotina diagnóstica ser realizada segundo uma série de parâmetros, além de nos instrumentar para modificá-las quando necessário. A aversão de muitos médicos em relação à matemática não se justifica diante da perda de sua capacidade de pensar os exames e as rotinas diagnósticas de forma crítica.

#### **Referências bibliográficas**

**Gianrossi R [et al.]** Exercise induced ST depression in the diagnosis of coronary artery disease. A meta-analysis. [Artigo] // Circulation. - 1989. - 1 : Vol. 80.

**Kwok Y [et al.]** Meta-analysis of exercise testing to detect coronary artery disease in women. [Artigo] // Am J Cardiology. - 1999. - 5 : Vol. 83.

**Medronho Roberto A e Pérez Maurício de Andrade** Testes Diagnósticos [Seção do Livro] // Epidemiologia / A. do livro Carvalho de [et al.]. - RJ : Atheneu, 2002.