

# AVALIAÇÃO

## DE TESTES DIAGNÓSTICOS

<b>INTRODUÇÃO</b>	
Reprodutibilidade ou repetibilidade .....	22
Validade ou acurácia .....	22
Relação entre precisão e acurácia .....	22
<b>REPRODUTIBILIDADE</b>	
Pesquisas de Laboratório, Pesquisa Clínica e Epidemiológica .....	23
Avaliação de Reprodutibilidade .....	23
Índice <i>Kappa</i> ( <i>k</i> ) .....	23
<b>VALIDADE DE UM TESTE DIAGNÓSTICO</b>	
Sensibilidade e Especificidade .....	25
Co-positividade e co-negatividade .....	26
Ponto de corte para delimitar resultados positivos .....	26
Valor preditivo do teste .....	27
Valor preditivo positivo .....	27
Valor preditivo negativo .....	27
Relação entre o valor preditivo e prevalência .....	27
<b>ERRO SISTEMÁTICO E ERRO ALEATÓRIO NA DETERMINAÇÃO DA SENSIBILIDADE E ESPECIFICIDADE</b>	
Erro Aleatório .....	27
Cálculo do tamanho da amostra para avaliar a sensibilidade e especificidade .....	28
Erro Sistemático .....	29
Viés de amostragem .....	29
Viés de mensuração .....	29
Viés de publicação .....	29
Princípios básicos para avaliar um teste diagnóstico/triagem .....	29
<b>ROTEIRO PARA DETERMINAR A VALIDADE DE UM TESTE</b> .....	<b>31</b>
<b>REFERÊNCIAS PARA LEITURA</b> .....	<b>32</b>
<b>EXERCÍCIOS</b> .....	<b>33</b>
<b>DICIONÁRIO DE BANCO DE DADOS</b> .....	<b>38</b>

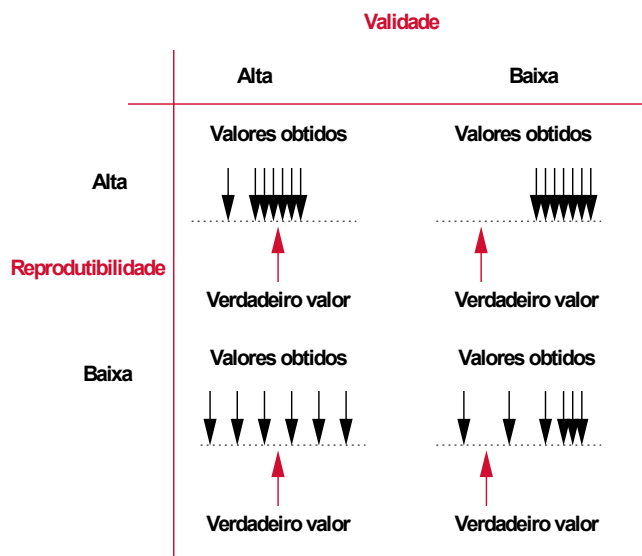
## INTRODUÇÃO

A avaliação da qualidade de testes diagnósticos é um tema de interesse da investigação clínica e epidemiológica. Em pesquisa epidemiológica, "testes diagnósticos" são entendidos não apenas como exames laboratoriais, mas, também, referem-se a procedimentos diversos como interrogatório clínico, exame físico e métodos propedêuticos diversos. O desempenho de um teste diagnóstico depende da ausência de desvios da verdade (ausência de viés) e da precisão (o mesmo teste aplicado ao mesmo paciente ou amostra deve produzir os mesmos resultados): respectivamente da validade e da reprodutibilidade do "teste". Neste módulo são discutidos dois conceitos básicos da qualidade de um teste diagnóstico: reprodutibilidade e validade e os aspectos relativos ao delineamento e análise destes estudos.

. **Reprodutibilidade ou repetibilidade** é a consistência de resultados quando o exame se repete. Por exemplo, dois radiologistas que lêem de forma independente as mesmas radiografias e chegam ao mesmo diagnóstico alcançam o nível máximo de reprodutibilidade. Mas, os dois especialistas podem estar igualmente corretos ou igualmente errados em seus diagnósticos.

. **Validade ou acurácia** refere-se ao grau em que o teste ou uma estimativa baseada em um teste é capaz de determinar o verdadeiro valor do que está sendo medido. A validade informa se os resultados representam a "verdade" ou o quanto se afastam dela. Por exemplo, o ECG é um teste de maior validade, comparado à auscultação cardíaca feita com o estetoscópio, no intuito de detectar alterações cardiovasculares típicas da doença de Chagas. Um teste "dip-stick" para detecção de antígeno utilizado para diagnóstico de malária por *P. falciparum* pode ter 100% de acurácia quando for capaz de produzir resultados positivos para todas as amostras de pacientes infectados e produzir resultados negativos para os indivíduos negativos.

. **Relação entre precisão e acurácia.** A Figura abaixo mostra a relação entre o valor verdadeiro de uma medida quantitativa e o valor obtido pelo estudo em termos de baixa e alta validade e reprodutibilidade. Com baixa reprodutibilidade e estando a média dos valores obtidos pelo estudo próxima do verdadeiro valor, o teste poderá ter validade, mas, mesmo assim, terá pouca utilidade. Por outro lado, uma alta repetibilidade da medida (resultados idênticos ou próximos quando o teste diagnóstico é repetido) não assegura validade pois os valores obtidos podem estar distantes do valor verdadeiro, ou seja, podem estar errados. Como esse aspecto é fundamental para separar corretamente doentes de saudáveis, a validade e a reprodutibilidade têm de ser adequadamente mensuradas, no sentido de avaliar a qualidade de um exame diagnóstico e, conseqüentemente, a informação por ele produzida. É importante aferir ambos os parâmetros, tanto com referência a novos testes introduzidos no mercado, como testes já em uso mas, aplicados em outros contextos.



Adaptado de Beaglehole et al, 1993

## REPRODUTIBILIDADE

. **Reprodutibilidade, repetibilidade ou precisão** - é a habilidade do teste em produzir resultados consistentes (quase os mesmos resultados) quando realizados independentemente e sob as mesmas condições. Por exemplo, um teste bioquímico é considerado de alta reprodutibilidade quando se obtém praticamente o mesmo resultado após várias testagens repetidas e de forma independente. Entretanto, se o aparelho eletrônico utilizado para realização do teste não estiver adequadamente calibrado, o teste pode ter alta reprodutibilidade, mas, produzir resultados consistentemente errados. O mesmo conceito de reprodutibilidade pode ser usado em situações mais gerais, como por exemplo, comparando-se os resultados de lâminas em diferentes ocasiões (variabilidade intra-observador).

### . Pesquisas de Laboratório, Pesquisa Clínica e Epidemiológica

Melhores resultados de reprodutibilidade são geralmente obtidos no trabalho de laboratório, onde as condições de operação podem ser mais controladas (um só observador, aparelhos de alta precisão, calibrados, com pouco uso, uso de amostras controle, ambiente livre de maiores perturbações e horário apropriado). Por outro lado, em pesquisas clínicas e epidemiológicas, raramente obtém-se o nível de reprodutibilidade encontrado em investigações de laboratório. O diagnóstico clínico, por exemplo, é um processo subjetivo, e, por isto, suscetível a interpretações discordantes, mesmo entre clínicos competentes e experientes. Em geral, um nível baixo de reprodutibilidade tende a atenuar as verdadeiras correlações entre eventos. Isto limita a utilidade do diagnóstico clínico em pesquisas populacionais, pois prejudica a investigação de associações entre fatores de risco e danos à saúde.

### . Avaliação de Reprodutibilidade

Há diversas maneiras de verificar a concordância de resultados entre leituras de um mesmo evento ou comparar métodos diagnósticos diferentes, e assim, estimar o erro cometido na sua aferição. Os resultados podem ser expressos sob forma de variável dicotômica (positivo / negativo), categórica (normal / anormal / níveis limítrofes), em medidas contínuas (miligramas, mililitros) ou títulos de sorologia. Este é um dos aspectos que influencia a forma de análise dos resultados. Geralmente, independente do tipo de dado produzido pelos testes diagnósticos, os médicos/epidemiologistas tendem a reduzi-lo à variáveis dicotômicas ou expressas em categorias para tornar a interpretação mais útil na prática. A comparação dos resultados pode ser apresentada através da taxa global de concordância entre os examinadores ou pelo indicador *Kappa*.

. **Índice Kappa ( $k$ )** - Uma maneira muito utilizada para expressar a confiabilidade de um teste é através do índice  $k$  que constitui um avanço em relação à taxa geral de concordância, por ser um indicador de concordância ajustada, pois leva em consideração, a concordância devida à chance. O  $k$  informa a proporção de concordância não aleatória (além da esperada pela chance) entre observadores ou medidas da mesma variável categórica, e seu valor varia de "menos 1" (completo desacordo) a "mais 1" (concordância total). Se a medida concorda mais frequentemente do que seria esperado pela chance, então o índice  $k$  é positivo; se a concordância é completa  $k = 1$ . Zero indica o mesmo que leituras feitas ao acaso. A Tabela 1 apresenta os valores do  $k$  e respectivas interpretações.

**Tabela 1- Escala de concordância do *Kappa***

<b>Kappa</b>	<b>Concordância</b>
<b>&lt;0,00</b>	<b>Nenhuma</b>
<b>0,00-0,20</b>	<b>Fraca</b>
<b>0,21-0,40</b>	<b>Sofrível</b>
<b>0,41-0,60</b>	<b>Regular</b>
<b>0,61-0,80</b>	<b>Boa</b>
<b>0,81-0,99</b>	<b>Ótima</b>
<b>1,00</b>	<b>Perfeita</b>

*Adaptado de Landis & Koch, Biometrics, 1977*

A Tabela 2 exemplifica o cálculo de *k*. Cento e vinte lâminas contendo esfregaços de gota espessa de sangue para pesquisa de hematozoários (malária) foram preparadas em condições uniformes e interpretadas por dois microscopistas independentes. O primeiro identificou 20 lâminas positivas e 100 negativas enquanto o segundo diagnosticou respectivamente 30 e 90 gerando 106 resultados concordantes (18+88) e 14 de discordantes (2+12). A taxa geral de concordância foi de 88,3% (106/120) e o valor de *k* = 65%.

**Tabela 2 – Concordância entre dois observadores nas leituras de laminas para pesquisa de hematozoários.**

	<b>Microscopista 1</b>		<b>Total</b>
	<b>(+)</b>	<b>(-)</b>	
<b>Microscopista 2</b>			
<b>(+)</b>	<b>18 (a)</b>	<b>12 (b)</b>	<b>30</b>
<b>(-)</b>	<b>2 (c)</b>	<b>88 (d)</b>	<b>90</b>
<b>Total</b>	<b>20</b>	<b>100</b>	<b>120</b>

O índice *kappa* é estimado como:

$$K = \frac{Po - Pe}{1 - Pe}$$

sendo:

Po = Proporção de concordâncias observadas

Pe = Proporção de concordâncias esperadas

$$Po = \frac{a + d}{a + b + c + d}$$

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

Para interpretação do  $k$  deve-se levar em conta:

**tipo de evento e outros fatores** - o nível de concordância depende do tipo de evento, fatores relacionados ao examinador, ao procedimento sendo testado e ao ambiente onde as observações são realizadas. Também, a diminuição do número de categorias de resultados (valores positivos e negativos ao invés de valor alto, médio, baixo e muito baixo) tende a aumentar a concordância.

**prevalência** - a prevalência do diagnóstico ou evento na população, afeta o resultado final. Baixas prevalências tendem à estar associadas a baixos níveis de reprodutibilidade, pois o valor de  $k$  depende da concordância devida ao acaso. É possível encontrar-se baixos níveis de reprodutibilidade, devido à baixa prevalência do evento e não à erros relacionados ao procedimento diagnóstico empregado. Por este motivo, deve-se informar a prevalência juntamente com os resultados do  $k$ .

**independência da avaliação** - as avaliações devem ser independentes umas das outras, princípio também aplicável à verificação da validade/accurácia. Isto significa que quando um examinador repete o teste, deve ignorar resultados prévios, obtidos por ele ou por outro examinador, para evitar a possibilidade de ser influenciado por este conhecimento e prejudicar a avaliação, mesmo involuntariamente.

## VALIDADE DE UM TESTE DIAGNÓSTICO

A validade de um teste refere-se à quanto, em termos quantitativos ou qualitativos, um teste é útil para diagnosticar um evento (validade simultânea ou concorrente) ou para predizê-lo (validade preditiva). Para determinar a validade, compara-se os resultados do teste com os de um padrão (padrão ouro): esse pode ser o verdadeiro estado do paciente, se a informação está disponível, um conjunto de exames julgados mais adequados, ou uma outra forma de diagnóstico que sirva de referência. O teste diagnóstico ideal deveria fornecer, sempre, a resposta correta, ou seja, um resultado positivo nos indivíduos com a doença e um resultado negativo nos indivíduos sem a doença. Além do que, deveria ser um teste rápido de ser executado, seguro, simples, inócuo, confiável e de baixo custo.

### . Sensibilidade e Especificidade

Para definir os conceitos de sensibilidade e especificidade, serão utilizados como exemplos, testes com resultados dicotômicos, isto é, resultados expressos em duas categorias: positivos ou negativos.

A Tabela 3 mostra as relações entre os resultados de um teste e o diagnóstico verdadeiro. O teste é considerado positivo (anormal) ou negativo (normal), e a doença presente ou ausente. Assim, na avaliação de um teste diagnóstico existem 4 interpretações possíveis para o resultado do teste: duas em que o teste está correto e duas em que está incorreto. O teste está correto quando ele é positivo na presença da doença (resultados verdadeiros positivos), ou negativo na ausência da doença (resultados verdadeiros negativos). Por outro lado, o teste está incorreto quando ele é positivo na ausência da doença (falso positivo), ou negativo quando a doença está presente (falso negativo). Os melhores testes diagnósticos são aqueles com poucos resultados falso-positivos e falso-negativos.

**Tabela 3. Validade de um Teste Diagnóstico**

		Doença (padrão-ouro)		
		Presente	Ausente	
Teste	Positivo	Verdadeiro positivo a	Falso-positivo b	a + b
	Negativo	Falso-negativo c	Verdadeiro negativo d	c + d
		a + c	b + d	N(a+b+c+d)

As seguintes proposições/Indicadores podem ser calculados da comparação dos resultados da tabela:

**Sensibilidade:**  $a/(a+c)$

**Especificidade:**  $d/(b+d)$

**Prevalência (real):**  $(a+c)/N$

**Prevalência estimada (teste):**  $(a+b)/N$

**Valor preditivo positivo:**  $a/(a+b)$

**Valor preditivo negativo:**  $d/(c+d)$

**Classificação correta (acurácia):**  $(a+d)/N$

**Classificação incorreta:**  $(b+c)/N$

. **Sensibilidade** - é a capacidade que o teste diagnóstico/triagem apresenta de detectar os indivíduos verdadeiramente positivos, ou seja, de diagnosticar corretamente os doentes.

. **Especificidade** - é a capacidade que o teste diagnóstico/triagem tem de detectar os verdadeiros negativos, isto é, de diagnosticar corretamente os indivíduos saudáveis.

. **Co-positividade e co-negatividade** - são termos utilizados em substituição, respectivamente, à sensibilidade e especificidade, quando o padrão empregado é outro teste considerado de referência para a doença em questão e não os diagnósticos de certeza de presença ou ausência de doença. São também designados sensibilidade relativa e especificidade relativa.

. **Ponto de corte para delimitar resultados positivos** - O teste ideal, com 100% de sensibilidade e especificidade raramente existe na prática, pois a tentativa de melhorar a sensibilidade frequentemente tem o efeito de diminuir a especificidade. Em algumas situações clínicas os resultados são obtidos através de variáveis contínuas, não havendo uma separação clara e inquestionável entre o que é "normal" e "anormal". Para a definição do ponto de corte de positividade o investigador deverá levar em conta a importância relativa da sensibilidade e especificidade do teste diagnóstico, ponderando sobre as implicações dos dois possíveis erros. Em indicações de procedimentos de risco (certas cirurgias), por exemplo, deve-se evitar resultados

falso-positivos; nestes casos, o ponto de corte deve ser definido de tal forma que aumente a especificidade do teste. Por outro lado, em triagens sorológicas em bancos de sangue para prevenção de transmissão de infecções nas quais a não detecção de casos acarretará risco para a população, o ponto de corte deverá ser estabelecido tendo como objetivo alcançar 100% de sensibilidade do teste para que não ocorram resultados falso-negativos, em que pese o aumento da proporção de falso-positivos. Para aumentar a sensibilidade em uma triagem pode-se utilizar mais do que um teste diagnóstico, em paralelo considerando-se como positivo as amostras que apresentarem pelo menos uma reação positiva. Em inquéritos populacionais, testes com alta sensibilidade devem ser utilizados, quando a prevalência da infecção na população em geral for baixa. Por outro lado, em clínica, é comum realizarem-se testes em série. Testes adicionais são realizados para confirmar resultados positivos ou negativos previamente obtidos.

. **Valor preditivo do teste** - No contexto epidemiológico e clínico, a validade de um marcador sorológico diz respeito à extensão com que ele pode prever a ocorrência da doença / infecção. Nessas circunstâncias, devemos estar preparados para responder à seguinte questão: dado que o teste apresentou resultado positivo (ou negativo), qual a probabilidade do indivíduo ser realmente doente (ou sadio)? Esse atributo do teste é conhecido como Valor Preditivo (VP) podendo ser positivo (VPP) ou negativo (VPN), e é determinado pela interação de três variáveis: a sensibilidade e a especificidade do teste e a prevalência da doença no grupo de estudo.

**Valor preditivo positivo** - é a proporção de doentes entre os positivos pelo teste. No exemplo da Tabela 2 teríamos 60% (18/30), o que equivale a dizer que em cada 10 testes positivos, 6 indivíduos seriam realmente doentes.

**Valor preditivo negativo** - é a proporção de sadios (sem a doença) entre os negativos ao teste. Ainda em relação à Tabela 2, teríamos um VPN de 98% (88/90); a cada 100 testes negativos, 98 seriam sadios.

#### . **Relação entre valor preditivo e prevalência**

Enquanto a sensibilidade e especificidade de um teste são propriedades inerentes ao teste e não variam a não ser por erro técnico, os VPs dependem da prevalência da doença na população de estudo. O VPP aumenta com a prevalência enquanto os VPN diminuem. Assim, quando a doença é rara o VPP é baixo, pois a maior parte dos exames positivos pertencem a sadios, representando resultados falso-positivos. Por outro lado, O VPN é alto em baixas prevalências. Os resultados falso-positivos e falso-negativos podem ser minimizados utilizando-se a combinação de testes, em paralelo (dois ou mais testes realizados simultaneamente) ou em série (dois ou mais testes realizados em seqüência), para a definição de resultado positivo. Se a intenção é reduzir resultados falso positivos (e aumentar a especificidade), um diagnóstico positivo deverá ser confirmado somente quando pelo menos dois testes diferentes forem positivos. Por outro lado, para reduzir resultados falso negativos (e aumentar a sensibilidade), um único teste positivo seria suficiente para considerar um diagnóstico positivo. Por exemplo, o teste será positivo se os 2 testes forem positivos, ou negativo se os 2 forem negativos.

### **ERRO SISTEMÁTICO E ERRO ALEATÓRIO NA DETERMINAÇÃO DA SENSIBILIDADE E ESPECIFICIDADE**

**Erro Aleatório** - Os estudos de avaliação de testes diagnósticos estão sujeitos a erros ao acaso; alguns pacientes com a doença apresentarão resultado normal do teste diagnóstico. Este tipo de erro pode ser avaliado calculando-se o intervalo de confiança para a sensibilidade e especificidade do novo teste. O intervalo de confiança indica o espectro de variação dos resultados obtidos para que se possa compará-los com os testes convencionais. Por exemplo, considere a comparação de dois testes A e B em uso na prática clínica; o teste A com sensibilidade de 80% e especificidade de 85%, calculadas após testagem em centenas de indivíduos e o novo teste diagnóstico (B) foi positivo em

10 indivíduos em um total de 10 pacientes com a doença, o que equivale a 100% de sensibilidade, e foi negativo em 9 de 10 indivíduos sem a doença (especificidade=90%). Apesar da sensibilidade e especificidade deste novo teste (b) serem maiores que as descritas para o teste (a), o intervalo de confiança de 95% tanto para a sensibilidade (61%-100%) como para a especificidade (55%-97%) mostrou um espectro grande de variação, com sobreposição destes intervalos com os do teste convencional; este fato decorre do pequeno número de indivíduos testados. Por este motivo, não é possível concluir que o novo teste (b) tenha um melhor desempenho do que o teste convencional (a).

Uma das estratégias para minimizar os erros aleatórios é estimar o tamanho da amostra para determinar a validade do teste diagnóstico, baseado na construção de intervalos de confiança, definindo-se espectros que incluam o valor que se deseja obter para a sensibilidade e especificidade do teste. Isto significa calcular dois tamanhos da amostra: um para a sensibilidade do teste e outro para a especificidade.

### **Cálculo do tamanho da amostra para avaliar a sensibilidade e especificidade**

O cálculo do tamanho da amostra, para variáveis dicotômicas, segue os mesmos princípios estabelecidos em estudos descritivos/estudos de prevalência (Quadro abaixo), sendo necessária as seguintes informações:

- (1) estimativa da proporção esperada da positividade na população (quando maior de 50% utilize a proporção de pessoas com resultados negativos)
- (2) amplitude do intervalo de confiança que se deseja
- (3) definição do intervalo de confiança (geralmente 95%)

$$N = Z * Z (P (1-P)) / (D * D) \quad \text{Onde:}$$

**P=** proporção esperada  
**D=** semi-amplitude do intervalo de confiança  
**Z =** 1,96 (para  $\alpha=0,05$  e IC 95%)

Por exemplo, em um estudo para determinar a sensibilidade de um novo teste diagnóstico para malária, espera-se que 80% dos pacientes com malária tenham teste positivo (resultado de estudo piloto). Quantos indivíduos com malária deverão ser testados para se estimar uma sensibilidade do teste de 80% com intervalo de 95% de confiança e precisão do teste de 0,04?. Considerando as 3 informações necessárias para o cálculo do tamanho da amostra, teríamos:

- (1) proporção esperada de casos com malária com teste positivo = 0,20 (80% é maior que 50%; portanto, a proporção de indivíduos com malária e teste negativo é 20%)
- (2) espectro do intervalo de confiança = 0,08 Utilize a semi-amplitude (0,04 acima ou 0,04 abaixo) como o erro máximo aceitável
- (3) intervalo de confiança = 95%



Utilizando-se a fórmula anexa, seria necessário aplicar o teste em 384 pacientes com malária para se estimar uma sensibilidade de 80% do teste com intervalo de 95% de confiança de 76%-84%:

$$n = 1,96^2 ( 0,20 (1-0,20) ) / (0,04^2)$$

n = 384 pessoas

Os mesmos procedimentos são válidos para o cálculo do tamanho da amostra para determinar a especificidade do teste. Por exemplo, se o investigador espera que 90% dos indivíduos sem malária tenham teste negativo, 216 indivíduos sem malária deveriam ser incluídos no estudo para determinação de uma especificidade de 90%  $\pm$ 0,04 com um intervalo de 95% de confiança.

**Erro Sistemático** - De forma geral, os estudos de testes diagnósticos estão sujeitos aos mesmos vieses que os estudos observacionais; os mais comuns são os vieses de amostragem, de medida do teste e de relato dos resultados.

. **Viés de amostragem** - neste tipo de erro, a amostra de estudo não é representativa da população alvo na qual o teste deverá ser utilizado. Por exemplo, a seleção de indivíduos provenientes de serviços de referência tende a incluir pessoas com formas graves da doença ou pacientes nos quais os testes sejam mais anormais do que seriam em outras formas clínicas da doença. Isto faz com que o estudo forneça resultados superestimados da sensibilidade do teste, diferindo da sensibilidade do teste em condições de rotina. Da mesma forma, o estudo fornecerá uma especificidade aumentada do teste, se indivíduos sem a doença forem selecionados como voluntários, pois estes tendem a ser mais saudáveis do que indivíduos com sintomas, recrutados de ambulatórios, porém sem a doença. A estratégia utilizada para minimizar este tipo de erro é selecionar amostras de população semelhante a qual o teste deverá ser utilizado. A escolha de amostras de populações nas quais a prevalência da doença é maior do que a habitualmente detectada, trará como consequência valores preditivos positivos superestimados. Uma situação muito comum é investigar um número igual de indivíduos com a doença e sem a doença, o que equivale a 50% de prevalência da doença. Para lidar com este viés o estudo deveria fornecer, também, resultados dos valores preditivos do teste ajustados para outras probabilidades de doença, para que o leitor possa avaliar a utilidade do teste de acordo com sua realidade na prática clínica.

. **Viés de mensuração** - Sempre que possível o investigador deve desconhecer quais indivíduos têm a doença e quais não têm, para evitar vícios de interpretação de resultados, especialmente nas situações limítrofes. Da mesma forma, o investigador deve permanecer mascarado em relação à realização dos testes diagnósticos. O ponto de corte deve ser definido antes da realização do teste.

. **Viés de publicação** - Existe uma tendência em se publicar somente os estudos que mostrem "sucesso" dos testes diagnósticos o que acarreta um bias de literatura. Para minimizar este viés, os estudos devem ser planejados com número suficiente de indivíduos para que os resultados tenham credibilidade e sejam devidamente divulgados.

### . **Princípios básicos para avaliar um teste diagnóstico/triagem**

O delineamento de estudos para avaliar/comparar a utilidade clínica ou populacional de testes diagnósticos deve incorporar dois aspectos. O primeiro deles diz respeito aos princípios da aleatorização e mascaramento. Se os pacientes são alocados aleatoriamente para receber o novo teste (versus aquele usado na rotina), os indivíduos que receberem este novo teste terão uma melhor evolução clínica? A comparação dos testes pode e deve ser feita nos mesmos indivíduos e amostras, para eliminar variações externas aos testes. O segundo aspecto a ser levado em conta refere-se à prática clínica vigente. O teste será aplicado nas mesmas condições de seu uso na clínica? O fato de um teste discriminar casos graves da doença não significa que será igualmente útil para distinguir pacientes portadores de doença leve dos demais pacientes com sintomas semelhantes.

Estudos conduzidos para determinar a validade de um teste diagnóstico apresentam estrutura semelhante aos estudos observacionais. Eles incluem a variável preditora (resultado do teste) e a variável de efeito (presença ou ausência da doença). A diferença entre eles reside nos seus objetivos. Na avaliação de testes diagnósticos, descreve-se a **intensidade da associação**, em termos de sensibilidade e especificidade (capacidade do teste em discriminar doentes de não doentes). Já, os estudos observacionais buscam determinar a presença de uma associação. Portanto, na análise da validade de um teste não basta apenas mostrar que existe uma associação entre o resultado do teste e a doença.

## ROTEIRO PARA DETERMINAR A VALIDADE DE UM TESTE

### ① *Certifique-se da necessidade do teste*

- . vantagens do novo teste em relação aos existentes
- . benefícios para pacientes, com a introdução do novo teste
- . custos para aplicação do teste em nível individual e em saúde pública

### ② *Estabeleça o critério de amostragem*

- . defina a população de referência e a população de estudo
- . esclareça a fonte de seleção dos participantes
- . informe sobre a inclusão de casos graves, moderados e leves

### ③ *Descreva o teste e o padrão de referência*

- . produto químico, imunobiológico, antígeno, anticorpo, procedência
- . etapas para processamento das reações
- . interpretação e categorização dos parâmetros a serem avaliados

### ④ *Descreva os procedimentos para a aplicação do teste e do padrão de referência*

- . aplicação dos testes de forma mascarada
- . codificação das amostras para envio ao laboratório - processamento sem conhecimento do status doente/não doente

### ⑤ *Calcule o tamanho da amostra*

- . estipule o número mínimo e suficiente de participantes para se estimar a sensibilidade e especificidade do teste com intervalo de 95% de confiança
- . estime o número de casos disponíveis ou a serem detectados no futuro próximo no local de seleção dos participantes

### ⑥ *Esclareça as questões éticas*

- . riscos da aplicação do teste e benefício da detecção de indivíduos positivos
- . atenção médica aos indivíduos positivos
- . confidencialidade dos resultados

### ⑦ *Análise de dados*

- . apresente os resultados em termos de sensibilidade, especificidade e valores preditivos com respectivos IC95%

*Adaptado de Hulley & Cummings, 1988*

## REFERÊNCIAS PARA LEITURA

BEAGLEHOLE, R., BONITA, R. & KJELLSTRÖM, T. *Basic Epidemiology*. World Health Organization, Geneva, 1993.

BUCK, A.A. & GART, J.J. Comparison of a screening test and a reference test in epidemiologic studies. I. Indices of agreement and their relation to prevalence. **American Journal of Epidemiology**,**83**:586-92, 1966.

FLEISS, J.L. *Statistical methods for rates and proportions*, 2<sup>nd</sup> ed. New York, John Wiley & Sons, 1981.

FLETCHER, R.M.; FLETCHER, S.W. & WAGNER, E.H. *Clinical Epidemiology, the essentials*. Baltimore - USA, Ed. Wawerly, 1983.

GALEN, R.S. & GAMBINO, S.R. *Beyond normality: the predictive value and efficiency of medical diagnosis*, New York:John Wiley & Sons ed., 1975.

HULLEY, S.B. & CUMMINGS, S.R. *Designing clinical research*. Williams & Wilkins Baltimore, 1988.

KRAEMER, H.C. & BLOCH, D.A. Kappa coefficients in epidemiology: an appraisal of a reappraisal. **Journal of Clinical**,**41**:959-68, 1988.

PEREIRA, M.G. *Epidemiologia: Teoria e Prática*. Rio de Janeiro. Guanabara Koogan, 1995.

## EXERCÍCIOS

**Arquivos:**

1. ecg.rec
2. malmt.rec
3. bloodch.rec

**1. Concordância entre leituras de eletrocardiogramas** - o arquivo **ecg.rec** listado em anexo contém os resultados de 100 Eletrocardiogramas que foram enviados para leitura de forma independente para dois centros.

**Questão 1.** Calcule a reprodutibilidade (Kappa) do diagnóstico de alteração eletrocardiográfica feito por 2 observadores (A e B). Interprete os resultados.

**Notas 1:**

- READ ECG.REC (para abrir o arquivo)
- TABLES CENTROA CENTROB
- anote os resultados
- tecle F9 para ir ao DOS
- digite EPITABLE para cálculo do Kappa (comando COMPARE / PROPORTIONS)
- utilize os dados da tabela anterior
- pressione F10 para sair do EPITABLE
- digite EXIT para voltar ao ANALYSIS
- CLOSE (para fechar o arquivo)

**2. Avaliação do diagnóstico parasitológico da malária** - o arquivo **malmt.rec** listado em anexo contém resultados da parasitemia de 141 indivíduos com febre, da demanda espontânea de um dia de atendimento de 2 Postos de Notificação de malária (Fundação Nacional de Saúde). Todas as lâminas foram encaminhadas de forma independente para o Centro de Controle de Qualidade (Ministério da Saúde) e os resultados comparados aos fornecidos pelos Postos de Notificação (serviços locais de saúde). A análise de dados foi planejada para avaliar a reprodutibilidade do diagnóstico de malária entre o Centro de Controle de Qualidade e os serviços locais de saúde e avaliar a sensibilidade destes serviços no diagnóstico de malária. Responda as questões abaixo utilizando o arquivo malmt.rec.

**Questão 2.** Qual a prevalência global (IC 95%) de malária de acordo com os resultados fornecidos pelo Centro de Controle de Qualidade? Entre os casos com malária qual a frequência por espécie?

**Notas 2:** READ MALMT.REC (para abrir o arquivo)  
FREQ LAMCQ /CI  
selecione apenas os indivíduos com malária  
SELECT LAMCQ=2  
FREQ DIAGCQ  
SELECT (para desativar a seleção)

**Questão 3.** Qual a prevalência global (IC 95%) de malária de acordo com os resultados fornecidos pelos Postos de Notificação da FNS? Qual a frequência de diagnóstico por espécie dentre os casos com malária?

**Notas 3:** FREQ LAMFNS /CI  
selecione apenas os indivíduos com malária  
SELECT LAMFNS=2  
FREQ DIAGFNS  
SELECT (para desativar a seleção)

**Questão 4.** Compare a proporção de malária por espécie utilizando os resultados fornecidos pelos Postos de Notificação (FNS) e o Centro de Controle de Qualidade (questões 2 e 3). Quais as implicações destes achados do ponto de vista do Programa de Controle da Malária?

**Questão 5.** Qual a reprodutibilidade (*Kappa*) do diagnóstico de malária entre os Postos de Notificação (FNS) e o Centro de Controle de Qualidade?

**Notas 5:** TABLES LAMFNS LAMCQ  
anote os resultados  
tecle F9 para ir ao DOS  
digite EPITABLE para cálculo do *Kappa* (comando  
COMPARE / PROPORTION)  
pressione ESC para voltar ao menu do EPITABLE

**Questão 6.** Qual a sensibilidade (IC 95%) dos Postos de Notificação (FNS) no diagnóstico da malária? Qual o total de casos falso negativos por espécie? Assuma como padrão de referência os resultados do Controle de Qualidade.

**Notas 6:** utilize os dados da tabela produzida na questão 5  
STUDY/ SCREENING  
pressione F10 para sair do EPITABLE  
digite EXIT para voltar ao ANALYSIS  
TABLES DIAGFNS DIAGCQ

**Questão 7.** Qual o valor preditivo positivo (IC 95%) da febre no diagnóstico de malária nestes Postos de Notificação? Interprete o custo-benefício da indicação do tratamento presuntivo de malária nesta área endêmica.

**Nota 7:**       FREQ LAMCQ  
                  CLOSE (para fechar o arquivo)

**Questão 8.** Discuta as implicações dos resultados desta investigação considerando a Estratégia Global de Controle da Malária - diagnóstico precoce e tratamento oportuno. Que estratégias você recomendaria para implementação da qualidade do diagnóstico de malária na área de estudo?

**3. Validação da sorologia para Doença de Chagas em bancos de sangue** - o arquivo **bloodch.rec** listado em anexo contém resultados de sorologia para infecção pelo *T.cruzi* de 1513 primodoadores de sangue triados nos 6 bancos de sangue da cidade de Goiânia (período 1988-1989) pelas técnicas de Hemaglutinação (HA) e Fixação de Complemento (FC). Amostras destes soros foram encaminhadas de forma independente a um dos Laboratórios de Referência para Doença de Chagas da OMS e os resultados comparados aos fornecidos pelos bancos de sangue (serviços locais de saúde). Detalhes da metodologia estão na referência **Andrade et al., 1992**. A análise teve como objetivo: (1) aferir a sensibilidade dos bancos de sangue na prevenção da transmissão transfusional da doença de Chagas e (2) avaliar a concordância do diagnóstico sorológico da infecção pelo *T.cruzi* em condições ideais (Laboratório de Referência) e na rotina dos bancos de sangue. Utilize o arquivo **bloodch.rec** para responder as questões abaixo.

**Questão 9.** Compare a prevalência de soropositividade ao *T.cruzi* pelas técnicas HA e IF realizadas pelo Laboratório de Referência. Avalie o benefício da utilização da triagem em paralelo comparada ao uso de apenas uma das técnicas.

**Notas 9:** READ BLOODCH.REC (para abrir o arquivo)  
criar as variáveis GRHA ( $HA \geq 16$ ) e GRIF ( $IF \geq 40$ )  
DEFINE GRHA \_\_ (grupo hemaglutinação)  
IF HA=-1 THEN GRHA="SS" (sem sorologia)  
IF HA>-1 AND HA<16 THEN GRHA="N" (negativo)  
IF HA>=16 THEN GRHA="R" (reativo)  
utilize o mesmo procedimento para criar a variável GRIF  
utilize somente os registros com resultado de GRHA  
SELECT GRHA <> "SS"  
FREQ GRHA  
SELECT (para desativar a seleção)  
utilize somente os registros com resultado de GRIF  
SELECT GRIF <> "SS"  
FREQ GRIF  
SELECT (para desativar a seleção)  
TABLES GRHA GRIF

**Questão 10.** Calcule a concordância (*Kappa*) do diagnóstico de infecção pelo *T.cruzi* por 2 testes distintos (HA e IF), realizados no laboratório de referência. Interprete os resultados.

**Notas 10:** utilize somente os pares disponíveis para ambas as técnicas  
SELECT GRHA <> "SS" and GRIF <> "SS"  
TABLES GRHA GRIF  
SELECT  
anote os resultados  
tecle F9 para ir ao DOS  
digite EPITABLE para cálculo do *Kappa* (comando COMPARE / PROPORTION)  
utilize os dados da tabela anterior  
pressione F10 para sair do EPITABLE  
digite EXIT para voltar ao ANALYSIS

**Questão 11.** Calcule a concordância do diagnóstico sorológico da infecção entre o laboratório de referência e os bancos de sangue. Para esta finalidade crie 2 novas variáveis com os resultados do laboratório de referência ("RES"=sorologia referência) e dos bancos de sangue ("POS"=resultado na triagem).



**Notas 11:** criar a variável RES (GRIF=R ou GRHA=R)  
DEFINE RES \_\_  
IF GRHA="SS" AND GRIF="SS" THEN RES="SS" ELSE RES="N"  
(recodificação para eliminar registros sem resultado)  
IF GRHA="R" OR GRIF="R" THEN RES="R"  
criar a variável POS (HABS=R ou FCBS=R)  
DEFINE POS \_\_  
IF HABS="SS" AND FCBS="SS" THEN POS="SS" ELSE POS="N"  
IF HABS="R" OR FCBS="R" THEN POS="R"  
utilize somente os pares disponíveis para ambos, laboratório de  
referência e bancos de sangue  
SELECT RES <> "SS" AND POS <> "SS"  
TABLES POS RES  
SELECT  
anote os resultados e organize a tabela  
tecle F9 para ir ao DOS  
digite EPITABLE para cálculo do Kappa (comando COMPARE /  
PROPORTION)  
utilize os dados da tabela anterior  
pressione F10 para sair do EPITABLE  
digite EXIT para voltar ao ANALYSIS

**Questão 12.** Calcule a sensibilidade e valores preditivos positivos e negativos do diagnóstico de infecção chagásica fornecido pelos Bancos de Sangue ("POS"). Para esta finalidade construa uma nova variável ("RESFIM") assumindo como padrão a positividade simultânea à HA e IF pelo Laboratório de Referência. Quantos indivíduos soropositivos deixariam de ser detectados pelos Bancos de Sangue na triagem de rotina de primodadores (falso negativos)? Dos doadores soropositivos encaminhados pelos Bancos de Sangue ao Ambulatório de Atenção ao Chagásico, qual o percentual de doadores que não teriam confirmação diagnóstica de doença de Chagas (responda considerando o VPP)?

**Notas 12:** criar a variável RESFIM (GRIF=R e GRHA=R)  
DEFINE RESFIM \_\_  
IF GRIF="R" AND GRHA="R" THEN RESFIM="R"  
IF GRIF="N" OR GRHA="N" THEN RESFIM="N"  
IF GRIF="SS" AND GRHA="SS" THEN RESFIM="SS"  
utilize somente os pares disponíveis para ambos, Laboratório  
de Referência e Bancos de Sangue  
SELECT POS <> "SS" AND RESFIM <> "SS"  
TABLES POS RESFIM  
SELECT  
anote os resultados  
tecle F9 para ir ao DOS  
utilizar EPITABLE para cálculo da sensibilidade (comando  
STUDY / SCREENING)  
pressione F10 para sair do EPITABLE  
digite EXIT para voltar ao ANALYSIS

**Questão 13.** Qual o número de indivíduos soropositivos detectados pela HA ("HABS") e pela FC ("FCBS") pelos Bancos de Sangue? Qual dos testes seria mais indicado para finalidade de triagem em Bancos de Sangue? Avalie o benefício da utilização da triagem em paralelo pelos Bancos de Sangue comparado ao uso de apenas uma técnica.

**Notas 13:** utilize somente os pares disponíveis para ambas as técnicas, HA e FC.  
SELECT HABS <> "SS" AND FCBS <> "SS"  
excluir BS6 pois os resultados estão disponíveis apenas para a HA  
SELECT BS<>6  
TABLES HABS FCBS  
SELECT (para desativar a seleção)  
CLOSE (para fechar o arquivo)

## REFERÊNCIAS BIBLIOGRÁFICAS

ANDRADE, A.L.S.S., MARTELLI, C.M.T., LUQUETTI, A.O., OLIVEIRA, O.S.O., SILVA, S.A. & ZICKER, F. Serologic screening for *Trypanosoma cruzi* among blood donors in Central Brazil. **Bulletin of the Pan American Health Organization**,26(2):157-164, 1992.

ANDRADE, A.L.S.S., MARTELLI, C.M.T., OLIVEIRA, R.M., ARIAS, J.R., ZICKER, F. & PANG, L. High prevalence of asymptomatic malaria in gold mining areas in Brazil. **Clinical Infectious Disease**,20(2): 475, 1995.

Arquivo: **ecg.rec**

Variável	Descrição	Código	Descrição do Código
<b>IDENTIFICA</b>	Número de identificação		
<b>CENTROA</b>	Leitura ECG pelo observador A	0 1	Sem alteração Com alteração
<b>CENTROB</b>	Leitura ECG pelo observador B	0 1	Sem alteração Com alteração

Arquivo: **malmt.rec**

Variável	Descrição	Código	Descrição do Código
<b>NO</b>	Número de identificação		
<b>LAMFNS</b>	Resultado Lâmina pela FNS	1 2	Negativo Positivo
<b>LAMCQ</b>	Resultado Lâmina pelo Controle Qualidade (MS)	1 2	Negativo Positivo
<b>DIAGCQ</b>	Diagnóstico pelo Controle Qualidade	1 2 3 4 5	<i>P.vivax</i> <i>P.falciparum</i> <i>P.malariae</i> mista Sem malária
<b>DIAGFNS</b>	Diagnóstico pela FNS	1 2 3 4 5	<i>P.vivax</i> <i>P.falciparum</i> <i>P.malariae</i> mista Sem malária

Arquivo: **bloodch.rec**

<b>Variável</b>	<b>Descrição</b>	<b>Código</b>	<b>Descrição do Código</b>
<b>NO</b>	Número de Identificação		
<b>BS</b>	Bancos de Sangue	1 2 3 4 5 6	Hemog Araújo Jorge Hemolabor Inst Hemoterapia Banco de Sangue Goiano Hospital das Clínicas
<b>HABS</b>	Hemaglutinação Banco de Sangue	R N SS	Reator Negativo Sem sorologia
<b>FCBS</b>	Fixação Complemento Banco de Sangue	R N SS	Reator Negativo Sem Sorologia
<b>HA</b>	Hemaglutinação Laboratório de Referência	-1 0 - 8 16 - 128	Sem Sorologia Negativo Positivo
<b>IF</b>	Imunofluorescência Laboratório de Referência	-1 0 40-1280	Sem Sorologia Negativo Positivo
<b>ELISA</b>	ELISA Laboratório de Referência	-1 0 - 0,9 1,0 - 3,5	Sem Sorologia Negativo Positivo