

## Medidas de acurácia de testes diagnósticos

# Testes diagnósticos

## O que são testes diagnósticos?

Ao fazer um diagnóstico, um médico estabelece um conjunto de diagnósticos alternativos com base nos sinais e sintomas do paciente. Progressivamente ele reduz suas alternativas até chegar à uma doença específica.

Alternativamente, ele pode ter fortes evidências de que o paciente tem uma determinada doença e deseja apenas sua confirmação. Para chegar à uma conclusão final o médico utiliza-se de testes diagnósticos:

- ▶ exames de laboratório (ex. dosagem de glicose)
- ▶ exame clínico (ex. auscultação do pulmão)
- ▶ questionário (ex. CDI (Children's Depression Inventory))

## O que são testes diagnósticos?

Um **teste diagnóstico** é um instrumento capaz de **diagnosticar** a doença com determinada **precisão**.

Os objetivos dos testes incluem a triagem de pacientes, o diagnóstico de doenças e o acompanhamento ou prognóstico referente a evolução do paciente.

Para cada teste diagnóstico existe um valor de referência que determina a classificação do resultado do teste como **negativo** ou **positivo**.

# Validade de testes diagnósticos

Um teste diagnóstico é considerado **útil** quando ele identifica bem a presença da doença.

**Validade/Acurácia** é a habilidade de um teste em separar os indivíduos doentes daqueles saudáveis.

Antes de ser adotado o teste deve ser avaliado para verificar sua capacidade de acerto.

Esta avaliação é feita aplicando-se o teste a dois grupos de pessoas: um grupo doente o outro não doente.

Nesta fase, o diagnóstico é feito por outro teste chamado **padrão ouro**.

# Testes padrão ouro



- ▶ **Referência** para detecção e/ou quantificação de determinado analito.
- ▶ Usado como padrão para comparação de novos testes que foram desenvolvidos e, assim, obter sua **validade clínico-diagnóstica**.
- ▶ Hipoteticamente o teste padrão ouro identifica **todos** os não doentes e **todos** os doentes.
- ▶ Na prática, o teste perfeito não existe.

# Avaliação de testes de diagnóstico

Esquema padrão dos dados para um teste diagnóstico

	$T_+$	$T_-$	Total
$D_+$	a	b	$n_1 = a + b$
$D_-$	c	d	$n_2 = c + d$
Total	$a + c$	$b + d$	$n = a + b + c + d$

►  $T_+$ : O resultado do teste é positivo

►  $T_-$ : O resultado do teste é negativo

►  $D_+$ : Indivíduo portador da doença

►  $D_-$ : Indivíduo não portador da doença

# Sensibilidade e especificidade

Na análise da **qualidade/acurácia** de testes diagnósticos, temos interesse em conhecer duas probabilidades condicionais

- ▶ **Sensibilidade:** probabilidade de o teste ser positivo, sabendo-se que o paciente que está sendo examinado é realmente doente, ou seja,

$$s = P(T_+ | D_+)$$

- ▶ **Especificidade:** probabilidade de o teste ser negativo sabendo-se que o paciente examinado não tem a doença, ou seja,

$$e = P(T_- | D_-)$$



## Estimação da sensibilidade e especificidade

Usando a definição de probabilidade condicional, a sensibilidade pode ser estimada por

$$s = P(T_+|D_+) = \frac{P(T_+ \cap D_+)}{P(D_+)} = \frac{a}{a + b}$$

e a especificidade, por

$$e = P(T_-|D_-) = \frac{P(T_- \cap D_-)}{P(D_-)} = \frac{d}{c + d}$$

## Estimação da sensibilidade e especificidade

As estimativas dos complementares da sensibilidade e da especificidade são mostrados na tabela abaixo

	$T_+$	$T_-$	Total
$D_+$	$s = \frac{a}{a+b}$	$1-s = \frac{b}{a+b}$	1
$D_-$	$1-e = \frac{c}{c+d}$	$e = \frac{d}{c+d}$	1
-			

## Cálculo da sensibilidade e especificidade

@Malacarne19 avaliaram o desempenho de testes para o diagnóstico de tuberculose pulmonar em populações indígenas no Brasil. Os resultados para o teste rápido molecular (TRM) em comparação à cultura de escarro (teste padrão ouro) para todas as amostras de escarro combinadas são mostrados na tabela abaixo.

Cultura do escarro	Teste Rápido Molecular (TRM)		Total
	Teste positivo (T <sub>+</sub> )	Teste negativo (T <sub>-</sub> )	
Com TB (D <sub>+</sub> )	54	4	58
Sem TB (D <sub>-</sub> )	7	401	408
Total	61	405	466

## Cálculo da sensibilidade e especificidade

A sensibilidade e a especificidade para o teste rápido molecular (TRM) são dadas, respectivamente, por

$$s = \frac{a}{a + b} = \frac{54}{54 + 4} = 0,9310$$

e,

$$e = \frac{d}{c + d} = \frac{401}{7 + 401} = 0,9828$$

# Combinação da sensibilidade e especificidade

Algumas medidas importantes podem ser obtidas pela **combinação** da sensibilidade e da especificidade:

- ▶ *Odds ratio*
- ▶ Índice de Youden
- ▶ Razão de verossimilhanças
- ▶ Acurácia

## Odds ratio

Em aulas anteriores, definimos a **chance** de um evento ocorrer ou *odds* como sendo

$$\frac{p}{1 - p}$$

em que  $p$  é a probabilidade do evento ocorrer. Isto é, a razão entre a probabilidade de um evento ocorrer e de não ocorrer.

A medida conhecida como *odds ratio* (*OR*) é definida como a razão de duas *odds*.

## Odds ratio

No contexto de qualidade de testes de diagnóstico, a medida  $OR$  é definida por:

$$OR = \frac{\frac{s}{1-s}}{\frac{e}{1-e}} = \frac{se}{(1-s)(1-e)}$$

A definição resulta da razão do produto da sensibilidade e especificidade pelo produto dos seus complementares.

# Odds ratio

Interpretação:

- ▶  $OR \approx 1$ : indica que a chance de um resultado de teste positivo é a mesma para pacientes com a condição ( $D_+$ ) e sem a condição ( $D_-$ )
- ▶  $OR > 1$ : indica que a chance de um resultado de teste positivo é mais alta para pacientes **com** a condição.
- ▶  $OR < 1$ : indica que a chance de um resultado de teste positivo é mais alta para pacientes **sem** a condição.



## Cálculo da odds ratio

Voltando ao exemplo do teste rápido molecular (TRM) para o diagnóstico de tuberculose pulmonar em populações indígenas

Cultura do escarro	Teste Rápido Molecular (TRM)		Total
	Teste positivo (T <sub>+</sub> )	Teste negativo (T <sub>-</sub> )	
Com TB (D <sub>+</sub> )	54	4	58
Sem TB (D <sub>-</sub> )	7	401	408
Total	61	405	466

## Cálculo da odds ratio

Baseado nos dados acima, vimos que

$$s = \frac{54}{54 + 4} = 0,9310 \text{ e } e = \frac{401}{7 + 401} = 0,9828$$

Assim,

$$OR = \frac{se}{(1-s)(1-e)} = \frac{0,9310 \times 0,9828}{(1 - 0,9310) \times (1 - 0,9828)} = 770,96$$

- **INTERPRETAÇÃO:** A chance de um resultado de teste positivo é de mais de 770 vezes em pacientes com tuberculose pulmonar em relação à do grupo sem a doença.

# Índice de Youden

O índice de Youden ( $J$ ) varia entre -1 e 1, e é definido como

$$J = s + e - 1$$

Interpretação:

- ▶ Para um teste perfeito,  $J = 1$
- ▶ Quanto mais altas as somas de classificações corretas do teste, nos casos em que a condição do paciente é conhecida, isto é  $s + e$ , maior será  $J$ .

## Valores referência para o Índice de Youden

Sensibilidade	Especificidade	Índice de Youden (J)	Implicações
1.00	1.00	1.00	Teste perfeito, fer
0.90	0.85	0.75	Excelente teste, al
0.80	0.90	0.70	Teste muito bom,
0.70	0.95	0.65	Bom teste, adequa
0.85	0.80	0.65	Bom teste, adequa
0.50	1.00	0.50	Teste moderado, a
0.60	0.70	0.30	Teste justo, recom
0.40	0.90	0.30	Teste justo, recom
0.30	0.80	0.10	Teste ruim, não re
0.20	0.50	-0.30	Teste inadequado,

## Cálculo do Índice de Youden

Voltando ao exemplo do teste rápido molecular (TRM) para o diagnóstico de tuberculose pulmonar em populações indígenas. Neste exemplo, vimos que  $s = 0,9310$  e  $e = 0,9828$ , de forma que

$$J = s + e - 1 = 0,9310 + 0,9828 - 1 = 0,9138$$

- **INTERPRETAÇÃO:** De acordo com o Índice de Youden, o teste rápido molecular (TRM) é um excelente teste, altamente confiável para o diagnóstico de tuberculose pulmonar em populações indígenas.

## Razão de verossimilhanças

O índice da razão de verossimilhanças ( $RV$ ) é definido como a razão entre probabilidades do resultado do teste (positivo ou negativo), entre aqueles com a condição ( $D_+$ ) e sem a condição ( $D_-$ ). Assim,

$$RV_+ = \frac{P(T_+|D_+)}{P(T_+|D_-)} = \frac{s}{1-e} \text{ e } RV_- = \frac{P(T_-|D_+)}{P(T_-|D_-)} = \frac{1-s}{e}$$

O índice  $RV$  indica a magnitude da evidência de que determinado resultado do teste (positivo ou negativo) fornece a favor da presença da condição em relação à sua ausência.

# Razão de verossimilhanças

Interpretação:

- ▶  $RV \approx 1$ : indica que o resultado do teste é igualmente provável nos grupos com e sem a condição.
- ▶  $RV > 1$ : indica que o resultado do teste é mais provável entre aqueles com a condição comparado aos sem a condição.
- ▶  $RV < 1$ : indica que o resultado do teste é mais provável entre aqueles sem a condição comparado aos com a condição.

## Cálculo da Razão de verossimilhanças

Voltando ao exemplo do teste rápido molecular (TRM) para o diagnóstico de tuberculose pulmonar em populações indígenas. Neste exemplo, vimos que  $s = 0,9310$  e  $e = 0,9828$ , de forma que

$$RV_+ = \frac{P(T_+|D_+)}{P(T_+|D_-)} = \frac{s}{1-e} = \frac{0,9310}{1-0,9828} = 54,13$$

e,

$$RV_- = \frac{P(T_-|D_+)}{P(T_-|D_-)} = \frac{1-s}{e} = \frac{1-0,9310}{0,9828} = 0,07$$



## Cálculo da Razão de verossimilhanças

- **INTERPRETAÇÃO:** Um resultado positivo é aproximadamente 54 vezes mais provável em pacientes com tuberculose pulmonar do que os não doentes e o resultado negativo é aproximadamente 14 vezes  $\left(\frac{1}{0,07} = 14,24\right)$  mais provável em pacientes não doentes.

## Acurácia

A acurácia ou probabilidade de um resultado de teste correto é definida por

$$acc = P(T_+ \cap D_+) + P(T_- \cap D_-)$$

que, pode ser pensada como a proporção de verdadeiros positivos ( $VP$ ) ou verdadeiros negativos ( $VN$ ).

Trata-se da média ponderada de sensibilidade e especificidade, com pesos iguais à prevalência ( $p$ ) e ao complementar da prevalência ( $1 - p$ ).

$$acc = P(T_+ \cap D_+) + P(T_- \cap D_-) = ps + (1 - p)e$$

Também podemos estimar a acurácia usando  $acc = \frac{a + d}{a + b + c + d}$

## Cálculo da acurácia

Voltando ao exemplo do teste rápido molecular (TRM) para o diagnóstico de tuberculose pulmonar em populações indígenas.

Cultura do escarro	Teste Rápido Molecular (TRM)		Total
	Teste positivo (T <sub>+</sub> )	Teste negativo (T <sub>-</sub> )	
Com TB (D <sub>+</sub> )	54	4	58
Sem TB (D <sub>-</sub> )	7	401	408
Total	61	405	466

Neste exemplo, vimos que  $s = 0,9310$  e  $e = 0,9828$ . A prevalência da doença é estimada por  $p = \frac{58}{466} = 0,12$ , de forma que

$$acc = ps + (1-p)e = 0,12 \times 0,9310 + (1-0,12) \times 0,9828 = 0,976$$

## Cálculo da acurácia

ou ainda,

$$acc = \frac{54 + 401}{466} = 0,976$$

- **INTERPRETAÇÃO:** Esperam-se 97,6% de resultados corretos ao aplicar o teste rápido molecular (TRM) para o diagnóstico de tuberculose pulmonar em populações indígenas.

## Valor de predição de um teste

Os índices de sensibilidade e especificidade são ilustrativos e bons sintetizadores das qualidades gerais de um teste mas têm uma limitação séria: **não ajudam a decisão da equipe médica que, recebendo um paciente com resultado positivo do teste, precisa avaliar se o paciente está ou não doente.**

Não se pode depender apenas da sensibilidade e da especificidade, pois estes índices são provenientes de uma situação em que há certeza total sobre o diagnóstico, o que não acontece no consultório médico.

Daí, temos a necessidade de dois outros índices que refletem melhor a realidade prática.

## Valor de predição de um teste

À equipe médica seria mais interessante conhecer o valor de predição positiva (VPP) e o valor de predição negativa (VPN) de um teste:

- ▶ **Valor da predição positiva (VPP):** probabilidade do paciente estar **realmente doente** quando o resultado do **teste é positivo**, ou seja,

$$VPP = P(D_+|T_+)$$

- ▶ **Valor da predição negativa (VPN):** probabilidade do paciente **não estar doente** quando o resultado do **teste é negativo**, ou seja,

$$VPN = P(D_-|T_-)$$

## Valor de predição de um teste

A maneira mais fácil de calcular o  $VPP$  e o  $VPN$  é através da tabela abaixo, sugerida por @Vecchio66.

Seja  $p = P(D_+)$  a **prevalência** (ou proporção) da doença na população de interesse, ou a **probabilidade de doença pré-teste**.

---

População	Proporção	Proporção com resultado	
		Positivo	Negativo
Doente	$p$	$ps$	$p(1-s)$
Sadia	$1 - p$	$(1-p)(1-e)$	$(1-p)e$
Total	$1$	$ps + (1-p)(1-e)$	$p(1-s) + (1-p)e$

---

## Valor preditivo positivo e negativo

O valor de predição positiva é obtido dividindo-se a frequência de pacientes doentes pelo total dos testes positivos

$$VPP = \frac{ps}{ps + (1 - p)(1 - e)}$$

De forma análoga, considerando-se os pacientes sadios obtemos o valor de predição negativa

$$VPN = \frac{(1 - p)e}{p(1 - s) + (1 - p)e}$$



## Valor preditivo positivo e negativo

Observações:

- ▶ Note que as expressões de  $VPP$  e  $VPN$  dependem do conhecimento da prevalência  $p$ , ou seja, da probabilidade de resultados corretos de diagnóstico.
- ▶ Já a sensibilidade e especificidade não dependem do conhecimento de  $p$ .

## Cálculo do VPP e VPN

Voltando ao exemplo do teste rápido molecular (TRM) para o diagnóstico de tuberculose pulmonar em populações indígenas.

Cultura do escarro	Teste Rápido Molecular (TRM)		Total
	Teste positivo (T <sub>+</sub> )	Teste negativo (T <sub>-</sub> )	
Com TB (D <sub>+</sub> )	54	4	58
Sem TB (D <sub>-</sub> )	7	401	408
Total	61	405	466

Neste exemplo, vimos que  $s = 0,9310$  e  $e = 0,9828$ . A prevalência da doença é estimada por  $p = \frac{58}{466} = 0,12$ , de forma que

## Cálculo do VPP e VPN

$$VPP = \frac{ps}{ps + (1-p)(1-e)} = \frac{0,12 \times 0,9310}{0,12 \times 0,9310 + (1-0,12) \times (1-0,9828)}$$

e,

$$VPN = \frac{(1-p)e}{p(1-s) + (1-p)e} = \frac{(1-0,12) \times 0,9828}{0,12 \times (1-0,9310) + (1-0,12) \times 0,9828}$$

## Cálculo do VPP e VPN

- ▶ **INTERPRETAÇÃO:** Note que os valores de  $VPP$  e  $VPN$  são bastante altos. Se o resultado do teste rápido molecular (TRM) é negativo, a probabilidade do paciente não ter tuberculose pulmonar é de 99%. Por outro lado, o fato de o teste diagnóstico dar positivo elevou a probabilidade de o paciente estar doente de 12% para 88%.
- ▶ Após a aplicação do teste rápido molecular (TRM), são esperados 12% de falso-positivos e 1% de falso-negativos.

# Cálculo do VPP e VPN

Considerações importantes:

- ▶ Quanto maior o valor da sensibilidade (mais próximo de 1), mais o valor do  $VPN$  vai se aproximando de 1.
  - ▶ Assim um teste **muito sensível** é útil para **descartar o diagnóstico** se ele der **negativo**.
- ▶ Quanto maior o valor de especificidade (mais próximo de 1), mais o valor do  $VPP$  vai se aproximando de 1.
  - ▶ Assim um teste **muito específico** é útil para **confirmar o diagnóstico** se ele der **positivo**.

## Decisões incorretas

Um **teste ideal** seria aquele que apresentasse 100% de sensibilidade e de especificidade. **Na prática** isso não acontece, de forma que podemos definir duas condições de **decisões incorretas** baseados em testes diagnósticos:

- ▶ **Probabilidade de Falso-Positivo ( $PFP$ ):** Proporção de indivíduos que não tem a doença entre aqueles com o teste positivo, ou seja,

$$PFP = P(D_-|T_+) = 1 - P(D_+|T_+) = 1 - VPP$$

- ▶ **Probabilidade de Falso-Negativo ( $PFN$ ):** Proporção de indivíduos que tem a doença entre aqueles com o teste negativo, ou seja,

$$PFN = P(D_+|T_-) = 1 - P(D_-|T_-) = 1 - VPN$$

## Decisões incorretas

Essas quantidades podem ser estimadas, respectivamente, por

$$PFP = 1 - VPP = 1 - \frac{ps}{ps + (1-p)(1-e)} = \frac{(1-p)(1-e)}{ps + (1-p)(1-e)}$$

e,

$$PFN = 1 - VPN = 1 - \frac{(1-p)e}{p(1-s) + (1-p)e} = \frac{p(1-s)}{p(1-s) + (1-p)e}$$

## Cálculo de PFP e PFN

Voltando ao exemplo do teste rápido molecular (TRM) para o diagnóstico de tuberculose pulmonar em populações indígenas.

Cultura do escarro	Teste Rápido Molecular (TRM)		Total
	Teste positivo (T <sub>+</sub> )	Teste negativo (T <sub>-</sub> )	
Com TB (D <sub>+</sub> )	54	4	58
Sem TB (D <sub>-</sub> )	7	401	408
Total	61	405	466



## Curva ROC (Receiver Operator Characteristic)

O valor escolhido como **ponto de corte** influenciará as características do teste.

Na curva 2, quanto menor o ponto de corte (ponto A), maior é a sensibilidade, mas menor é a especificidade. Neste caso, se pretendemos um **teste muito sensível e menos específico**, escolhe-se este ponto para obter uma **menor proporção de falsos negativos** e uma **maior proporção de falsos positivos**.

O ponto de corte localizado no ponto B permite pôr em evidência os valores para os quais existe maior optimização da sensibilidade em função da especificidade.

Já quanto maior o ponto de corte (ponto C), maior especificidade, porém, menor sensibilidade. Neste caso, se pretendemos um **teste muito específico e menos sensível**, escolhe-se o ponto C, obtendo-se uma **menor proporção de falsos positivos** e uma **maior proporção de falsos negativos**.

## Curva ROC (Receiver Operator Characteristic)

O valor do **ponto de corte** é definido como um valor que pode ser selecionado, arbitrariamente, pelo pesquisador entre os valores possíveis para a variável de decisão.

Acima do ponto de corte, o paciente é classificado como positivo (teste positivo, presença de doença) e, abaixo, é classificado como negativo (teste negativo, ausência de doença).

Para cada ponto de corte são calculados valores de sensibilidade e especificidade, dispostos no gráfico.

A aplicação Curva ROC (@freire21) mostra o efeito sobre a sensibilidade e especificidade quando variamos o ponto de corte de uma variável de teste contínua.

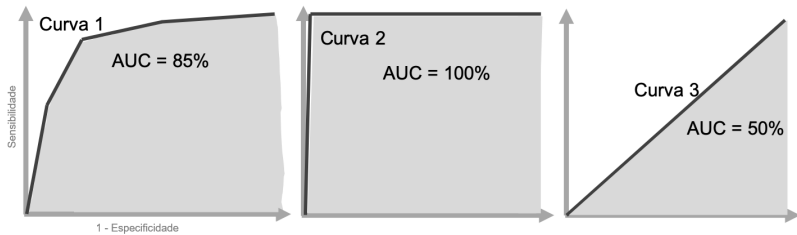
# Área sob a curva ROC (AUC-ROC)

Existem algumas formas de **resumir a acurácia** de um teste diagnóstico

A **área sob a curva ROC (AUC-ROC)** é uma medida resumo usual do desempenho de um teste, já que ela é estimada levando-se em consideração **todos** os possíveis valores para sensibilidade e especificidade.

Essa área pode variar de 0 a 1. Quanto mais próxima a curva ROC do canto superior esquerdo do gráfico, mais próxima de 1 é a área sob a curva e melhor é o desempenho do teste.

# Área sob a curva ROC (AUC-ROC)



## Escolha entre testes diagnósticos: mais sensibilidade ou mais especificidade?

Idealmente, os testes utilizados devem ter alta sensibilidade e especificidade.

Na prática, nem sempre existem testes disponíveis com **características ótimas**.

É relativamente comum a situação em que há mais de um teste disponível. Daí, surge a necessidade da comparação entre eles para a escolha mais adequada.

## Escolha entre testes diagnósticos: mais sensibilidade ou mais especificidade?

Suponha que dois testes diagnósticos estão disponíveis: um com alta sensibilidade, mas relativamente baixa especificidade, e outro com alta especificidade e relativamente baixa sensibilidade.

Qual seria a melhor escolha?

## Escolha entre testes diagnósticos: mais sensibilidade ou mais especificidade?

Os valores de predição ( $VPP$  e  $VPN$ ) dependem conjuntamente da sensibilidade ( $s$ ), da especificidade ( $e$ ) e da prevalência ( $p$ ).

- ▶ Quando a prevalência é baixa, o  $VPP$  é mais influenciado pela especificidade.
- ▶ O  $VPN$  é pouco influenciado tanto pela sensibilidade quanto pela especificidade e é alto.
- ▶ Para prevalência alta, o  $VPP$  é próximo de um, independentemente dos valores da sensibilidade e da especificidade.
- ▶ O  $VPN$  é influenciado mais pela sensibilidade do que pela especificidade.

## Escolha entre testes diagnósticos: mais sensibilidade ou mais especificidade?

Assim, existem evidências para as seguintes conclusões:

- 1) Um **teste com alta especificidade** deve ser usado quando a **prevalência da doença é relativamente baixa** (*doença rara*), mesmo que o teste tenha relativamente baixa sensibilidade.
- 2) Um **teste com alta sensibilidade** deve ser usado quando a **prevalência da doença é alta** (*doença comum*), mesmo que o teste tenha relativamente baixa especificidade.

Neste exemplo, vimos que  $s = 0,9310$  e  $e = 0,9828$ . A

prevalência da doença é estimada por  $p = \frac{58}{466} = 0,12$ , de forma que



## Cálculo de PFP e PFN

$$PFP = \frac{(1-p)(1-e)}{ps + (1-p)(1-e)} = \frac{(1-0,12) \times (1-0,9828)}{0,12 \times 0,9310 + (1-0,12) \times (1-0,9828)}$$

e,

$$PFN = \frac{p(1-s)}{p(1-s) + (1-p)e} = \frac{0,12 \times (1-0,9310)}{0,12 \times (1-0,9310) + (1-0,12) \times 0,0172}$$

- **INTERPRETAÇÃO:** Esperam-se 12% de resultados falsos positivos ao aplicar o teste rápido molecular (TRM) para o diagnóstico de tuberculose pulmonar em populações indígenas. Por outro lado, o TRM mostra-se útil para excluir a doença, uma vez que a probabilidade de um resultado falso-negativo é de 1%.

## Combinação de testes diagnósticos

Muitas vezes, para o diagnóstico de certa doença, dispomos apenas de testes com *VPN* ou *VPP* baixos ou se existe um bom teste, este é **muito caro** ou oferece **grande risco e/ou desconforto** ao paciente.

Nestas circunstâncias, uma opção é o uso de uma **combinação de testes mais simples**. A associação de testes **eleva a qualidade** do diagnóstico, diminuindo o número de **resultados incorretos**.

Quando dois ou mais testes são usados para se chegar a um diagnóstico é preciso saber como são obtidos os índices de qualidade do testes múltiplos. Restringiremos ao caso de apenas dois testes e as idéias podem ser estendidas para o caso de mais testes.

## Formas de combinação de testes

As maneiras mais simples de se formar um **teste múltiplo** a partir dos resultados de dois testes são os esquemas em **paralelo (simultâneo)** e em **série (sequencial)**.

No caso do **teste em paralelo**, se **um dos dois testes for positivo** o teste conjunto também é **positivo**.

O **teste em série** é considerado **positivo** apenas se **os dois testes individuais forem positivos**.

## Formas de combinação de testes

Chamando os testes originais de  $A$  e  $B$ , o teste em paralelo de  $T_p$  e o em série de  $T_s$ , e usando a linguagem de conjuntos temos que:

$$T_{p+} = A_+ \cup B_+$$

e,

$$T_{s+} = A_+ \cap B_+$$

As sensibilidades e especificidades de  $T_p$  e  $T_s$  são calculadas com o auxílio das regras de cálculo de probabilidades de eventos vistas anteriormente.

## Combinação em paralelo

É de maior utilidade em casos de emergência, quando se necessita de uma abordagem rápida, ou nos casos em que os pacientes são provenientes de lugares distantes.

A tabela abaixo mostra o resultado do teste em paralelo dependendo da classificação dos testes individuais A e B

---

Teste A	Teste B	Teste em paralelo
-	-	-
-	+	+
+	-	+
+	+	+

---

## Combinação em paralelo

Admitindo-se que os testes  $A$  e  $B$  são independentes, a sensibilidade do teste em paralelo ( $s_p$ ) é calculada por:

$$\begin{aligned}s_p &= P(T_{p+}|D_+) = P(A_+ \cup B_+|D_+) \\ &= P(A_+|D_+) + P(B_+|D_+) - P(A_+ \cap B_+|D_+)\end{aligned}$$

ou seja, a sensibilidade do teste em paralelo ( $s_p$ ) é dada por:

$$s_p = s_A + s_B - s_A \times s_B$$

## Combinação em paralelo

Admitindo-se que os testes  $A$  e  $B$  são independentes, a especificidade do teste em paralelo ( $e_P$ ) é calculada por:

$$\begin{aligned}e_p &= P(T_{p-}|D_-) = P(A_- \cap B_-|D_-) \\ &= P(A_-|D_-) \times P(B_-|D_-)\end{aligned}$$

ou seja, a sensibilidade do teste em paralelo ( $e_p$ ) é dada por:

$$e_p = e_A \times e_B$$

Os índices  $VPP$  e  $VPN$  são calculados utilizando-se a sensibilidade e especificidade da combinação em paralelo e a prevalência da doença na população de interesse.

## Combinação em série

Nesse caso, os testes são aplicados **consecutivamente**, sendo o **segundo teste** aplicado apenas se o primeiro apresentar **resultado positivo**. Assim, o **custo** desse tipo de combinação é **menor**.

Esse procedimento é indicado em situações onde não há necessidade de rápido atendimento e quando o paciente poder ser acompanhado ao longo do tempo.



## Combinação em série

A tabela abaixo mostra o resultado do teste em série dependendo da classificação dos testes individuais A e B

---

Teste A	Teste B	Teste em série
-	Desnecessário	-
+	-	-
+	+	+

---

Se os dois testes A e B são independentes, a sensibilidade ( $s_s$ ) deste teste é:

$$s_s = P(T_{s+}|D_+) = P(A_+ \cap B_+ | D_+) = P(A_+ | D_+) \times P(B_+ | D_+) = s_A \times s_B$$

## Combinação em série

A partir de raciocínio análogo, obtemos a expressão para a especificidade ( $e_s$ ):

$$\begin{aligned}e_s &= P(T_{s-}|D_-) = P(A_- \cup B_-|D_-) \\&= P(A_-|D_-) + P(B_-|D_-) - P(A_- \cap B_-|D_-)\end{aligned}$$

ou seja,

$$e_s = e_A + e_B - e_A \times e_B$$

## Combinação em série

Novamente os valores de  $VPP$  e  $VPN$  são obtidos usando-se a sensibilidade e especificidade calculadas acima.

Para os cálculos da sensibilidade e especificidade da associação em série e em paralelo, a **independência dos dois testes é crucial**. Quando os testes **não forem independentes**, não há uma forma analítica simples para se obter tais índices para um teste composto.

## Exemplo de combinação de testes diagnósticos

Imagine um paciente idoso com apresentando dores persistentes nas costas e no abdômen, e perda de peso. Na ausência de uma explicação para estes sintomas, a possibilidade de câncer de pâncreas é frequentemente levantada para se verificar esta possibilidade diagnóstica que ambos os testes de ultrassom e tomografia computadorizada do pâncreas sejam solicitados. A sensibilidade e especificidade de ambos os testes são dadas abaixo:

---

Teste	Sensibilidade	Especificidade
A: Ultrassom	0,80	0,60
B: Tomografia	0,90	0,90

---

## Exemplo de combinação de testes diagnósticos

Admitindo-se que os resultados dos dois testes sejam independentes, a sensibilidade e a especificidade dos testes combinados em paralelo seriam dadas, respectivamente, por

$$s_p = s_A + s_B - s_A \times s_B = 0,80 + 0,90 - 0,80 \times 0,90 = 0,98$$

e,

$$e_p = e_A \times e_B = 0,60 \times 0,90 = 0,54$$

Note que quando executados em paralelo, a sensibilidade combinada é maior que o mais sensível dos testes, mas a especificidade é menor.

## Exemplo de combinação de testes diagnósticos

Admitindo-se que os resultados dos dois testes sejam independentes, a sensibilidade e a especificidade dos testes combinados em série seriam dadas, respectivamente, por

$$s_s = s_A \times s_B = 0,80 \times 0,90 = 0,72$$

e,

$$e_s = e_A + e_B - e_A \times e_B = 0,60 + 0,90 - 0,60 \times 0,90 = 0,96$$

Note que quando executados em série, a especificidade combinada é maior que o mais específico dos testes, mas a sensibilidade é menor.

## Exemplo de combinação de testes diagnósticos

Admitindo-se que a prevalência do câncer de pâncreas no Brasil seja de 2%, encontre os  $VPP$ ,  $VPN$ ,  $PFP$  e  $PFN$  para cada um dos testes individuais e combinados.

Para o teste A: Ultrassom, temos

$$VPP = \frac{ps}{ps + (1 - p)(1 - e)} = \frac{0,02 \times 0,80}{0,02 \times 0,80 + (1 - 0,02) \times (1 - 0,60)}$$

e,

$$VPN = \frac{(1 - p)e}{p(1 - s) + (1 - p)e} = \frac{(1 - 0,02) \times 0,60}{0,02 \times (1 - 0,80) + (1 - 0,02) \times 0,60}$$

## Exemplo de combinação de testes diagnósticos

Para o teste B: Tomografia, temos

$$VPP = \frac{ps}{ps + (1-p)(1-e)} = \frac{0,02 \times 0,90}{0,02 \times 0,90 + (1-0,02) \times (1-0,90)}$$

e,

$$VPN = \frac{(1-p)e}{p(1-s) + (1-p)e} = \frac{(1-0,02) \times 0,90}{0,02 \times (1-0,90) + (1-0,02) \times 0,90}$$



## Exemplo de combinação de testes diagnósticos

Para os testes combinados em paralelo, temos

$$VPP = \frac{ps}{ps + (1-p)(1-e)} = \frac{0,02 \times 0,98}{0,02 \times 0,98 + (1-0,02) \times (1-0,54)}$$

e,

$$VPN = \frac{(1-p)e}{p(1-s) + (1-p)e} = \frac{(1-0,02) \times 0,54}{0,02 \times (1-0,98) + (1-0,02) \times 0,54}$$

## Exemplo de combinação de testes diagnósticos

Para os testes combinados em série, temos

$$VPP = \frac{ps}{ps + (1-p)(1-e)} = \frac{0,02 \times 0,72}{0,02 \times 0,72 + (1-0,02) \times (1-0,96)}$$

e,

$$VPN = \frac{(1-p)e}{p(1-s) + (1-p)e} = \frac{(1-0,02) \times 0,96}{0,02 \times (1-0,72) + (1-0,02) \times 0,96}$$

## Exemplo de combinação de testes diagnósticos

Resumindo os resultados em uma tabela,

---

Teste	Sensibilidade	Especificidade	VPP	VPN	PFP =
A: Ultrassom	0,80	0,60	0,0392	0,9932	0
B: Tomografia	0,90	0,90	0,1552	0,9977	0
Paralelo	0,98	0,54	0,0416	0,9992	0
Série	0,72	0,96	0,2686	0,9941	0

---

Interpretação:

- A combinação em paralelo apresenta alta sensibilidade (98%), mas sua especificidade é mais baixa que a dos testes isolados. Já a combinação em série apresenta alta especificidade (96%), enquanto que a sensibilidade é relativamente baixa comparada com os testes isolados.

## Exemplo de combinação de testes diagnósticos

- ▶ Como esperado, os valores de predição negativa ( $VPN$ ) são altos, tanto para os testes isolados como para os testes combinados. Entretanto, os valores de predição positiva ( $VPP$ ) não são altos, sendo que a combinação em série proporcionou o melhor resultado (26,86%)
- ▶ Por outro lado, a aplicação do teste combinado em paralelo mostra-se útil para excluir a doença, uma vez que a probabilidade de um resultado falso-negativo é de menos de 1%.

# Necessidade da combinação de testes

Há pelo menos duas situações em que a necessidade de combinação de testes surge naturalmente: **triagem** e **diagnóstico individual**.

- ▶ **Triagem:** É um tipo de procedimento que visa classificar pessoas assintomáticas quanto à probabilidade de terem ou não a doença.
  - ▶ É aplicado em grande número de pessoas de uma população
  - ▶ Não faz por si só o diagnóstico, mas aponta as pessoas que tem maior probabilidade de estarem doentes.
  - ▶ Para essas, é indicado um outro teste diagnóstico para comprovar ou não a presença da doença (testes em série).

# Necessidade da combinação de testes

- ▶ **Diagnóstico individual:** Aparecem os dois tipos de combinação: em série e em paralelo.
  - ▶ A combinação em paralelo é usada em casos de urgência ou para pacientes residentes em lugares distantes.
  - ▶ A combinação em série é usada em consultórios e clínicas hospitalares e em casos de testes caros e arriscados.

# Testes diagnósticos baseados em variáveis contínuas

Até agora, definimos a qualidade de um teste diagnóstico utilizando variáveis dicotômicas (teste positivo/teste negativo, doente/sadio)

Em muitas situações, variáveis quantitativas servem como indicador de uma doença:

- ▶ níveis de pressão diastólica para diagnóstico de hipertensão;
- ▶ glicemia elevada para diagnóstico de diabetes;
- ▶ entre outras.

# Testes diagnósticos baseados em variáveis contínuas

Para dados de variáveis contínuas, surge a questão de como **dicotomizar** de forma que os dados sejam colocados no formato da tabela considerada anteriormente.

A ideia é estabelecer o limite entre o que será considerado sadio ou doente, ou seja, o valor a partir do qual o teste é considerado positivo ou negativo.

A esse limite chamamos *valor de referência* ou *ponto de corte*.

Qual o melhor ponto de corte?

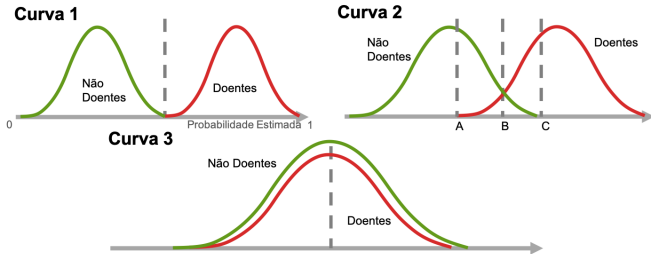
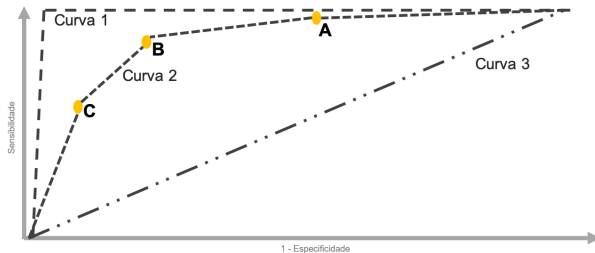


# Curva ROC (Receiver Operator Characteristic)

A **curva ROC** avalia o desempenho dos testes e a variação da sensibilidade e especificidade para diferentes valores de corte num gráfico que apresenta a proporção dos verdadeiros positivos (sensibilidade), no eixo Y, versus a proporção dos falsos positivos (1-especificidade), no eixo X.

Desta forma, a Curva ROC é um **gráfico de sensibilidade**.

# Curva ROC (Receiver Operator Characteristic)



## Curva ROC (Receiver Operator Characteristic)

- ▶ Observe que na curva 1, é representado o ponto de maior acurácia do teste (100%). Aqui temos a situação do teste ideal ( $s = 1$ ,  $e = 1$ ).
- ▶ A curva 2 representa uma boa curva, por exemplo, uma acurácia de 85%, que significa que em 85% das vezes em que for utilizado, o teste discriminará os verdadeiros positivos e negativos, mas também significa que em 15% das vezes fornecerá um falso resultado.
- ▶ A curva 3, expressa a acurácia de 50% e significa que o teste é tão bom quanto qualquer escolha ao acaso. Em geral, uma acurácia maior ou igual a 70% é considerada satisfatória.

## Curva ROC (Receiver Operator Characteristic)

O valor escolhido como **ponto de corte** influenciará as características do teste.

Na curva 2, quanto menor o ponto de corte (ponto A), maior é a sensibilidade, mas menor é a especificidade. Neste caso, se pretendemos um **teste muito sensível e menos específico**, escolhe-se este ponto para obter uma **menor proporção de falsos negativos** e uma **maior proporção de falsos positivos**.

O ponto de corte localizado no ponto B permite pôr em evidência os valores para os quais existe maior otimização da sensibilidade em função da especificidade.

Já quanto maior o ponto de corte (ponto C), maior especificidade, porém, menor sensibilidade. Neste caso, se pretendemos um **teste muito específico e menos sensível**, escolhe-se o ponto C, obtendo-se uma **menor proporção de falsos positivos** e uma **maior proporção de falsos negativos**.

# Curva ROC (Receiver Operator Characteristic)

O valor do **ponto de corte** é definido como um valor que pode ser selecionado, arbitrariamente, pelo pesquisador entre os valores possíveis para a variável de decisão.

Acima do ponto de corte, o paciente é classificado como positivo (teste positivo, presença de doença) e, abaixo, é classificado como negativo (teste negativo, ausência de doença).

Para cada ponto de corte são calculados valores de sensibilidade e especificidade, dispostos no gráfico.

A aplicação Curva ROC (@freire21) mostra o efeito sobre a sensibilidade e especificidade quando variamos o ponto de corte de uma variável de teste contínua.

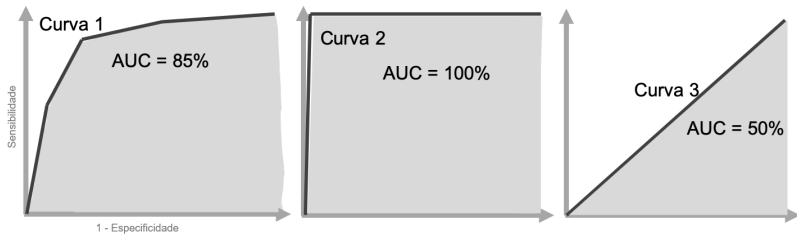
# Área sob a curva ROC (AUC-ROC)

Existem algumas formas de **resumir a acurácia** de um teste diagnóstico

A **área sob a curva ROC (AUC-ROC)** é uma medida resumo usual do desempenho de um teste, já que ela é estimada levando-se em consideração **todos** os possíveis valores para sensibilidade e especificidade.

Essa área pode variar de 0 a 1. Quanto mais próxima a curva ROC do canto superior esquerdo do gráfico, mais próxima de 1 é a área sob a curva e melhor é o desempenho do teste.

# Área sob a curva ROC (AUC-ROC)



## Escolha entre testes diagnósticos: mais sensibilidade ou mais especificidade?

Idealmente, os testes utilizados devem ter alta sensibilidade e especificidade.

Na prática, nem sempre existem testes disponíveis com **características ótimas**.

É relativamente comum a situação em que há mais de um teste disponível. Daí, surge a necessidade da comparação entre eles para a escolha mais adequada.



## Escolha entre testes diagnósticos: mais sensibilidade ou mais especificidade?

Suponha que dois testes diagnósticos estão disponíveis: um com alta sensibilidade, mas relativamente baixa especificidade, e outro com alta especificidade e relativamente baixa sensibilidade.

Qual seria a melhor escolha?

## Escolha entre testes diagnósticos: mais sensibilidade ou mais especificidade?

Os valores de predição ( $VPP$  e  $VPN$ ) dependem conjuntamente da sensibilidade ( $s$ ), da especificidade ( $e$ ) e da prevalência ( $p$ ).

- ▶ Quando a prevalência é baixa, o  $VPP$  é mais influenciado pela especificidade.
- ▶ O  $VPN$  é pouco influenciado tanto pela sensibilidade quanto pela especificidade e é alto.
- ▶ Para prevalência alta, o  $VPP$  é próximo de um, independentemente dos valores da sensibilidade e da especificidade.
- ▶ O  $VPN$  é influenciado mais pela sensibilidade do que pela especificidade.

## Escolha entre testes diagnósticos: mais sensibilidade ou mais especificidade?

Assim, existem evidências para as seguintes conclusões:

- 1) Um **teste com alta especificidade** deve ser usado quando a **prevalência da doença é relativamente baixa** (*doença rara*), mesmo que o teste tenha relativamente baixa sensibilidade.
- 2) Um **teste com alta sensibilidade** deve ser usado quando a **prevalência da doença é alta** (*doença comum*), mesmo que o teste tenha relativamente baixa especificidade.