



## SEÇÃO I

### MÉTODOS EM PSICOFÍSICA

- 1 Decisões clínicas e a teoria da detecção de sinal  
*J. Landeira-Fernandez, Sergio S. Fukusima*
- 2 Informações complementares sobre a teoria da detecção de sinal aplicada à psicofísica  
*Sérgio S. Fukusima, J. Landeira-Fernandez*
- 3 A utilização de métodos psicofísicos na medição da sensibilidade ao contraste visual em humanos  
*Nelson Torro Alves, Natanael Antonio dos Santos, Maria Lúcia de Bustamante Simas*
- 4 A medição de filtros de frequência espacial de banda estreita utilizando métodos psicofísicos  
*Renata Maria Toscano Barreto Lira Nogueira, Aline Mendes Lacerda, Maria Lúcia de Bustamante Simas, Natanael Antonio dos Santos*
- 5 Imagética motora: estudos cronométricos e aplicações clínicas  
*Maria Luíza Sales Rangel, Allan Pablo Lameira, Roberto Sena Fraga Filho, Luísa Azevedo Damasceno, Jessica Sanches Braga Figueira, Alice Andrade Pinho, Rayssa Carvalho de Souza, Fernanda Jazenko, Luiz G. Gawryszewski*



# Decisões clínicas e a teoria da detecção de sinal

*J. Landeira-Fernandez  
Sergio S. Fukusima*

## INTRODUÇÃO

O diagnóstico consiste no ato de inferir a presença ou a ausência de uma condição clínica. Como se trata de um processo inferencial, todo diagnóstico está sujeito a erros. Uma situação extremamente simples pode ilustrar esse processo decisório. Uma mulher, com o seu período atrasado, levanta a possibilidade de estar grávida. Para sanar essa dúvida, nada mais natural do que comprar um teste de gravidez e mergulhar a fita dentro de um pequeno frasco contendo uma amostra da urina. Dependendo da cor que essa fita apresente, é possível chegar, em menos de cinco minutos, a um diagnóstico acerca da presença ou ausência da gravidez.

Embora o resultado desse procedimento diagnóstico simples e barato seja relativamente seguro, deve-se reconhecer que erros podem ocorrer. Nenhum teste de gravidez pode garantir 100% de acerto. Discrepâncias podem existir entre o diagnóstico positivo ou negativo e o fato da mulher estar ou não grávida, fato esse que só se saberá muito tempo depois. Assim, é possível, que o resultado do teste de gravidez tenha sido positivo, mas na realidade a mulher não esteja grávida. Esse erro, denominado falso-positivo, pode ocorrer, por exemplo, nos casos em que a mulher tenha uma disfunção hormonal ou esteja tomando certos medicamentos que contenham o hormônio da gravidez (hormônio gonadotrofina coriônica humana). Por outro lado, o resultado do teste de gravidez pode ter sido negativo, mas a mulher pode estar, de fato, grávida. Esse erro, denominado falso-negativo, pode ocorrer, por exemplo, quando o teste é realizado de forma incorreta, principalmente se não for com a primeira urina da manhã, ou se o teste for feito muito prematuramente, logo no início da ausência da menstruação habitual.

A situação descrita ilustra o raciocínio lógico envolvido no processo decisório para se chegar a um diagnóstico. Obviamente, quando existem inúmeras evidências clínicas, todas convergindo para uma mesma direção, o processo decisório é relativamente trivial. Entretanto, a experiência clínica mostra que na maioria das vezes essas evidências são ambíguas ou incompletas. Mais ainda, quanto mais cedo for feito o diagnóstico, melhor o prognóstico. Nesse caso, o desenvolvimento de procedimentos que permitam chegar a um diagnóstico precoce de um quadro clínico certamente apresenta grande vantagem no planejamento de intervenções terapêuticas. É exatamente nesse contexto que técnicas relacionadas com a teoria da detecção de sinal podem ser extremamente úteis.

O ponto de partida para o emprego dessas técnicas é o de reconhecer a distinção entre duas condições completamente diversas: 1) a real existência ou não da condição clínica e; 2) a tomada de decisão (inferência) acerca da existência ou não dessa condição. A Tabela 1 apresenta uma combinação dessas duas condições, produzindo uma matriz de contingência 2 x 2.

Tabela 1 Matriz de contingência 2 x 2 que conjuga a existência (presença) ou não (ausência) de determinada condição clínica e o processo decisório envolvido na inferência que permite chegar a um diagnóstico acerca da existência (positivo) ou não (negativo) dessa condição

Matriz 2 x 2		Condição clínica (padrão-ouro)	
		Presente	Ausente
Diagnóstico (inferência)	Positivo	Verdadeiro-positivo (VP)	Falso-positivo (FP)
	Negativo	Falso-negativo (FN)	Verdadeiro-negativo (VN)

As duas colunas dessa matriz representam o fato da condição clínica existir ou não. Essas duas condições são de difícil acesso. Por essa razão, é necessário um longo e minucioso processo de avaliação que permita chegar à plena certeza da condição clínica. A esse procedimento denomina-se padrão-ouro. As duas linhas dessa matriz, por sua vez, representam os dois possíveis resultados de um procedimento diagnóstico que permite inferir a presença (diagnóstico positivo) ou a ausência (diagnóstico negativo) da condição clínica.

Conforme se pode observar na Tabela 1, a combinação dessas duas colunas com as duas linhas geram quatro células. Duas delas – verdadeiro-positivo (VP) e verdadeiro-negativo (VN) – indicam uma convergência entre a condição real do quadro clínico e a inferência diagnóstica que se faz acerca desse quadro. Portanto, essas duas células representam duas condições de acerto. A célula VP expressa o número de pessoas que de fato apresentam uma condição clínica e que o procedimento diagnóstico foi capaz de inferir corretamente a presença dessa condição. A célula VN indica o número de pessoas que de fato não apresentam a condição clínica e que o procedimento diagnóstico foi capaz de inferir corretamente a ausência de tal condição.

As duas outras células – falso-positivo (FP) e falso-negativo (FN) – indicam divergências entre a condição real do quadro clínico e a inferência que se faz acerca desse quadro. Portanto, essas duas outras células representam duas condições de erro. A célula FP expressa o número de pessoas que de fato não apresentam uma condição clínica, mas o procedimento diagnóstico inferiu erroneamente a presença de tal condição. Finalmente, a célula FN indica o número de pessoas que de fato apresentam uma condição clínica, mas o procedimento diagnóstico inferiu erroneamente a ausência dessa condição. É exatamente a partir desse raciocínio que a teoria da detecção de sinal pode auxiliar no processo de avaliação de um procedimento diagnóstico e assim amparar uma decisão clínica.

#### PARÂMETROS RELACIONADOS À QUALIDADE DO PROCEDIMENTO DIAGNÓSTICO: SENSIBILIDADE E ESPECIFICIDADE

O objetivo de qualquer procedimento diagnóstico é o de inferir a real existência ou não de uma condição clínica de forma clara e cada vez mais precoce. Por essa razão, é extremamente importante desenvolver e avaliar continuamente novos procedimentos diagnósticos que permitam atingir tais objetivos. Para que se possa realizar essa tarefa é necessário saber se de fato o paciente apresenta ou não a condição clínica na qual o novo procedimento está sendo avaliado. A princípio, pode-se imaginar por que se quer desenvolver um novo procedimento diagnóstico se já se sabe a condição clínica do paciente. A razão está relacionada com o fato de que para que se possa ter plena certeza de um diagnóstico são necessárias inúmeras informações que demandam tempo e recursos financeiros. Procedimentos diagnósticos mais simples que não prejudiquem a capacidade de inferir a real condição do quadro clínico do paciente apresentam vantagens óbvias. Sabe-se se o paciente apresenta ou não uma condição clínica com certeza muito tempo depois do início desse quadro. Nesse caso, intervenções terapêuticas terão pouco ou nenhum valor, dado o estágio avançado da condição clínica. De fato, a completa certeza de algumas patologias, via de regra, só podem ocorrer com procedimentos altamente invasivos, tais como cirurgias, biópsias ou mesmo procedimentos *post mortem*, como no caso de autópsias. Esses procedimentos altamente sofisticados que permitem ter total certeza em relação à condição clínica do paciente são denominados padrão-ouro, como já mencionado. Ele é fundamental para que se possa desenvolver procedimentos diagnósticos mais simples e mais baratos que permitam inferir a presença ou a ausência de um quadro clínico ainda em estágios iniciais com a mesma qualidade do padrão-ouro.

Sensibilidade e especificidade representam os dois parâmetros utilizados para avaliar a qualidade de um procedimento diagnóstico. Define-se sensibilidade como a capacidade do procedimento diagnóstico em revelar a presença de uma condição clínica quando de fato ela existe. De forma complementar, define-se especificidade como a capacidade do procedimento diagnóstico em revelar a ausência de uma condição clínica quando de fato ela não existe. A definição desses dois parâmetros pode ser representada por meio de matriz de contingência 2 x 2, conforme apresentado na Tabela 2.

Tabela 2 Estimativas da sensibilidade e da especificidade de um procedimento diagnóstico

Matriz 2 x 2		Condição clínica (padrão-ouro)	
		Presente	Ausente
Diagnóstico (inferência)	Positivo	Verdadeiro-positivo (VP)	Falso-positivo (FP)
	Negativo	Falso-negativo (FN)	Verdadeiro-negativo (VN)
Qualidade do diagnóstico		Sensibilidade $VP/(VP + FN)$	Especificidade $VN/(VN + FP)$

Conforme se pode observar na Tabela 2, a sensibilidade de um procedimento diagnóstico pode ser estimada por meio da razão entre o número de casos que foram corretamente diagnosticados de forma positiva (VP) e o número total de pessoas que de fato apresentam a condição clínica, independente dessa condição ter sido (VP) ou não erroneamente (FN) detectada pelo procedimento diagnóstico. Conforme representado na primeira coluna da matriz 2 x 2 da Tabela 2, pode-se expressar o cálculo da sensibilidade de um procedimento diagnóstico por meio da razão  $VP/(VP + FN)$ .

A especificidade de um procedimento diagnóstico, por sua vez, pode ser calculada por meio da razão do número de casos que foram corretamente diagnosticados de forma negativa (VN) e o número total de pessoas que de fato não apresentavam a condição clínica, independente de essa condição não ter sido (VN) ou ter sido erroneamente (FP) detectada pelo procedimento diagnóstico. Conforme representado na segunda coluna da matriz 2 x 2 da Tabela 2, pode-se expressar o cálculo da especificidade de um diagnóstico pela razão  $VN/(VN + FP)$ .

Notar que sensibilidade e especificidade são parâmetros que avaliam a qualidade de um diagnóstico. Quanto maior o valor da sensibilidade de um procedimento diagnóstico, maior a certeza que esse procedimento terá em revelar a presença de uma condição clínica quando de fato ela existe. Da mesma forma, quanto maior o valor da especificidade de um procedimento diagnóstico, maior a certeza de que esse procedimento terá a capacidade de revelar a ausência de uma condição clínica quando de fato ela não existe. Na verdade, a sensibilidade de um procedimento diagnóstico representa a taxa de VP, enquanto a especificidade representa a taxa de VN. Em valores percentuais, um procedimento diagnóstico ideal deve apresentar valores de sensibilidade e especificidade próximos de 100%.

#### PARÂMETROS RELACIONADOS À UTILIDADE CLÍNICA DO DIAGNÓSTICO: VALOR PREDITIVO POSITIVO E VALOR PREDITIVO NEGATIVO

A matriz de contingência 2 x 2 permite, também, derivar dois outros parâmetros associados a um procedimento diagnóstico. Nesse caso, leva-se em consideração as duas linhas da matriz 2 x 2 e não as duas colunas, como no caso da sensibilidade e da especificidade,

discutido anteriormente. A Tabela 3 expressa esses dois outros parâmetros denominados valor preditivo positivo e valor preditivo negativo.

Tabela 3 Estimativas dos valores preditivos positivo e negativo de um procedimento diagnóstico

Matriz 2 x 2	Condição clínica (padrão-ouro)		Utilidade do diagnóstico	
	Presente	Ausente		
Diagnóstico (inferência)	Positivo	Verdadeiro-positivo (VP)	Falso-positivo (FP)	Valor preditivo positivo $VP/(VP + FP)$
	Negativo	Falso-negativo (FN)	Verdadeiro-negativo (VN)	Valor preditivo negativo $VN/(VN + FN)$

A expressão “valor preditivo positivo” de um diagnóstico indica a probabilidade de um paciente com um diagnóstico positivo em relação a uma condição clínica de fato apresentar tal condição. Conforme se pode observar na Tabela 3, esse parâmetro pode ser estimado pela razão entre o número de casos que foram corretamente diagnosticados de forma positiva (VP) e o número total de diagnósticos positivos independente dessa condição clínica encontrar-se presente (VP) ou ausente (FP). Assim, pode-se expressar o cálculo do valor preditivo positivo de um procedimento diagnóstico por meio da razão  $VP/(VP + FP)$ . Esse parâmetro encontra-se representado na primeira linha da matriz 2 x 2 da Tabela 3.

Já a expressão “valor preditivo negativo” de um diagnóstico indica a probabilidade de um paciente com um diagnóstico negativo em relação a uma condição clínica de fato não apresentar tal condição. Conforme se observa na Tabela 3, esse parâmetro pode ser estimado pela razão entre o número de casos que foram corretamente diagnosticados de forma negativa (VN) e o número total de diagnósticos negativos independente dessa condição clínica encontrar-se ausente (VN) ou presente (FN). Assim, pode-se expressar o cálculo do valor preditivo negativo de um procedimento diagnóstico por meio da razão  $VN/(VN + FN)$ . Esse parâmetro encontra-se representado na primeira linha da matriz 2 x 2 da Tabela 3.

Os valores preditivos positivo e negativo possuem relevância clínica. Quanto maior o valor preditivo positivo de um teste, maior a certeza de que um paciente diagnosticado positivamente em relação a uma condição clínica de fato apresente tal condição. Da mesma forma, quanto maior o valor preditivo negativo de um teste, maior a certeza de que um paciente diagnosticado negativamente a uma condição clínica de fato não apresente tal condição. Em valores percentuais, espera-se que ambos os valores preditivos de um procedimento diagnóstico sejam próximos de 100%.

## ALGUNS EXEMPLOS

Até aqui foi discutido apenas aspectos abstratos acerca do emprego da teoria da detecção de sinal no processo de avaliação da qualidade e utilidade de um procedimento diagnóstico. Como ilustração para esses parâmetros, seguem alguns exemplos modificados de

uma série de trabalhos realizados para avaliar uma bateria neuropsicológica breve capaz de inferir prejuízos cognitivos a pacientes que sofrem de esclerose múltipla,<sup>1,2</sup> cujos valores foram ajustados de acordo com exemplos sugeridos por Loong.<sup>3</sup>

Suponha-se uma amostra com 100 pacientes com esclerose múltipla, sabendo-se que 30 deles de fato desenvolveram problemas cognitivos. Sabe-se disso graças a extensas avaliações neuropsicológicas e acompanhamentos clínicos por um longo período (padrão-ouro). Durante as primeiras consultas, quando ainda não se sabia se o paciente com esclerose múltipla apresentava ou não problemas cognitivos, aplicou-se a bateria neuropsicológica breve (BNB). No final desse estudo fictício, encontraram-se os resultados representados na Tabela 4.

Tabela 4 Resultados fictícios de um estudo realizado com 100 pacientes com esclerose múltipla a fim de investigar a capacidade de uma bateria neuropsicológica breve em inferir déficits cognitivos nesses pacientes. Neste exemplo, a prevalência do déficit cognitivo é de 30% (i.e., 30 entre os 100 pacientes com esclerose múltipla de fato apresentaram problemas cognitivos). A tabela também indica a sensibilidade e a especificidade, bem como os valores preditivos positivos e negativos desse procedimento diagnóstico

Matriz 2 x 2		Condição clínica (padrão-ouro)		Total de diagnósticos	Utilidade do diagnóstico
		Presente	Ausente		
Diagnóstico (inferência)	Positivo	24	14	38	Valor preditivo positivo 63,2%
	Negativo	6	56	62	Valor preditivo negativo 90,3%
Total da condição clínica		30	70	100	
Qualidade do diagnóstico		Sensibilidade 80%	Especificidade 80%		Prevalência 30%

Como se pode observar na Tabela 4, a BNB foi capaz de identificar corretamente 24 entre em 30 pacientes que de fato apresentaram problemas cognitivos. Seis diagnósticos negativos falharam. Dessa forma, pode-se afirmar que a sensibilidade desse procedimento diagnóstico é de 80%. Essa mesma bateria neuropsicológica identificou corretamente 56 entre 70 pacientes que de fato não apresentam problemas cognitivos. Quatorze diagnósticos positivos falharam. Conseqüentemente, pode-se afirmar que a especificidade dessa bateria é de 80%. Esses resultados indicam que a BNB apresenta valores de sensibilidade e especificidade bem razoáveis, uma vez que ambos os valores estão próximos de 100%.

Nesse mesmo estudo, é possível calcular ainda os valores preditivos positivo e negativo desse procedimento diagnóstico. Nesse caso, a BNB corretamente diagnosticou positivamente 24 casos entre todos os 38 diagnósticos positivos realizados. Quatorze diagnósticos positivos falharam. Assim, o valor preditivo positivo da BNB foi de 63,2%. Essa mesma BNB corretamente diagnosticou negativamente 56 casos entre todos os 62 diagnósticos negativos realizados (seis diagnósticos negativos falharam). Portanto, o valor preditivo negativo desse procedimento é de 90,3%. Esses resultados indicam que a BNB apresenta maior utilidade clínica para diagnósticos negativos em relação a diagnósticos positivos, uma vez que o valor

preditivo negativo desse procedimento encontra-se bem próximo de 100% (90,3%), sendo também maior que o seu valor preditivo positivo (63,2%).

Um aspecto importante acerca dos valores preditivos positivos e negativos é o fato de que esses dois parâmetros não dependem das características intrínsecas do procedimento diagnóstico que está sendo avaliado. Eles estão sujeitos à prevalência da condição clínica na população. Quanto maior a prevalência dessa condição clínica na população maior será o valor preditivo positivo e menor o valor preditivo negativo desse procedimento diagnóstico. O contrário também é verdadeiro. Por exemplo, pode-se assumir que a prevalência do déficit cognitivo entre os pacientes que sofrem de esclerose múltipla tenha subido para 60%. Caso a mesma BNB seja aplicada da mesma forma que no exemplo anterior, seria possível obter os resultados representados na Tabela 5.

Tabela 5 Resultados fictícios de estudo similar àquele apresentado na Tabela 4. Neste exemplo, a prevalência do déficit cognitivo é de 60%. Notar que os valores da sensibilidade e da especificidade permanecem os mesmos. Entretanto, o valor preditivo positivo subiu enquanto o valor preditivo negativo caiu em comparação ao exemplo apresentado na Tabela 4, na qual a prevalência do déficit cognitivo era de 30%

Matriz 2 x 2		Condição clínica (padrão-ouro)		Total de diagnósticos	Utilidade do diagnóstico
		Presente	Ausente		
Diagnóstico (inferência)	Positivo	48	8	56	Valor preditivo positivo 85,7%
	Negativo	12	32	44	Valor preditivo negativo 72,7%
Total da condição clínica		60	40	100	
Qualidade do diagnóstico		Sensibilidade de 80%	Especificidade 80%		Prevalência 60%

Observar que no caso da Tabela 5, em que a prevalência do déficit cognitivo subiu para 60%, os valores da sensibilidade e da especificidade desse procedimento diagnóstico permaneceram os mesmos (sensibilidade =  $48/60 = 80\%$ ; especificidade =  $32/40 = 80\%$ ). Entretanto, o valor preditivo positivo desse procedimento subiu para 86,7% ( $48/56$ ), enquanto o valor preditivo negativo caiu para 72,7% ( $32/44$ ). Portanto, o aumento da prevalência do déficit cognitivo entre pacientes com esclerose múltipla levou ao aumento do valor preditivo positivo desse procedimento diagnóstico, assim como à redução do seu valor preditivo negativo. Nesse caso, a utilidade clínica dessa bateria neuropsicológica passou a ser maior para diagnósticos positivos em relação ao exemplo anterior, em que a prevalência do déficit cognitivo era de apenas 30%. Nesse caso, o valor preditivo positivo encontra-se mais próximo de 100% (85,7%), além de ser maior que o valor preditivo negativo (72,7%).

A situação inversa encontra-se ilustrada na Tabela 6. Neste caso, a prevalência é de apenas 10%. Mais uma vez, os valores da sensibilidade e da especificidade desse procedimento diagnóstico não sofreram nenhuma alteração (sensibilidade =  $48/60 = 80\%$ ; especificidade =  $32/40 = 80\%$ ). Entretanto, o valor preditivo positivo desse procedimento caiu para 30,8%



(8/26), enquanto o valor preditivo negativo subiu para 97,3% (72/74). Portanto, a redução da prevalência do déficit cognitivo entre pacientes com esclerose múltipla levou à queda do valor preditivo positivo desse procedimento diagnóstico, assim como ao aumento do seu valor preditivo negativo. Da mesma forma que no exemplo anterior, esses resultados fazem sentido clínico, uma vez que a redução da prevalência do déficit cognitivo para 10% dá maior ênfase ao diagnóstico negativo. O valor preditivo negativo, além de estar próximo de 100% (97,3%), é bem maior que o valor preditivo positivo (30,8%).

Tabela 6 Resultados fictícios de estudo similar àquele apresentado na Tabela 4. Neste exemplo, a prevalência do déficit cognitivo é de apenas 10%. Notar que os valores da sensibilidade e da especificidade permanecem os mesmos, mas o valor preditivo positivo caiu enquanto o valor preditivo negativo subiu em comparação com o exemplo apresentado na Tabela 5, na qual a prevalência do comprometimento cognitivo era de 60%

Matriz 2 × 2		Condição clínica (padrão-ouro)		Total de diagnósticos	Utilidade do diagnóstico
		Presente	Ausente		
Diagnóstico (inferência)	Positivo	8	18	26	Valor preditivo positivo 30,8%
	Negativo	2	72	74	Valor preditivo negativo 97,3%
Total da condição clínica		10	90	100	
Qualidade do diagnóstico		Sensibilidade 80%	Especificidade 80%		Prevalência 10%

## EM BUSCA DO PONTO DE CORTE IDEAL: A CURVA ROC

Sensibilidade e especificidade são qualidades intrínsecas de um procedimento diagnóstico. Quando o resultado desse procedimento apresenta apenas duas categorias que definem a presença (diagnóstico positivo) ou a ausência (diagnóstico negativo), então esse procedimento diagnóstico apresenta apenas um único par de valores referentes à sensibilidade e à especificidade. Entretanto, a maioria dos procedimentos diagnósticos produz resultado sob a forma de variável contínua. Neste caso, os valores de sensibilidade e especificidade desse procedimento diagnóstico variarão de acordo com o ponto de corte empregado para que se possa chegar a um diagnóstico.

Sensibilidade e especificidade apresentam uma relação inversamente proporcional. Como esses dois parâmetros variam de acordo com diferentes pontos de corte, mudando-se o ponto de corte de um procedimento diagnóstico pode-se aumentar a sensibilidade desse procedimento em detrimento de sua especificidade e vice-versa. Calcado nessa relação inversamente proporcional entre sensibilidade e especificidade, o ponto de corte ideal é aquele que busca o equilíbrio entre sensibilidade e especificidade de um procedimento diagnóstico.

Pode-se ilustrar essa relação com o mesmo exemplo que utiliza uma BNB para detectar possíveis déficits cognitivos em pacientes que sofrem de esclerose múltipla. Em mais um estudo fictício, 100 pacientes com esclerose múltipla foram testados inicialmente com a

BNB. Muito tempo depois, várias baterias neuropsicológicas foram aplicadas para verificar aqueles que realmente apresentaram ou não comprometimento cognitivo. Por meio desse padrão-ouro, conclui-se que 40 desses pacientes apresentaram comprometimento cognitivo (i.e., prevalência de 40%).

A BNB gera resultados de números inteiros, cujos valores podem variar de 0 a 10. Quanto maior o valor desse resultado, maior o comprometimento cognitivo do paciente. A Tabela 7 apresenta os diferentes pontos de cortes calcados nesses possíveis resultados variando de 0 a 10, com os respectivos números de pacientes nas condições VP, FN, FP e VN para cada um desses pontos de corte. A tabela também apresenta os cálculos das respectivas sensibilidade e especificidade, bem como as médias desses dois parâmetros para cada ponto de corte.

Tabela 7 Número de pacientes nas condições VP, FN, FP e VN de acordo com o resultado em uma BNB utilizando diferentes pontos de corte. Para cada ponto de corte calculou-se a respectiva sensibilidade, especificidade, assim como a média desses dois parâmetros. Notar que para pontos de corte baixos (critério lenientes) a sensibilidade é alta enquanto a especificidade é baixa. Conforme o ponto de corte se torna mais rígido ocorre a inversão desses parâmetros. O ponto de corte ideal é aquele que busca o equilíbrio entre os valores de sensibilidade e especificidade (maior média entre esses dois parâmetros), que neste exemplo representa o ponto de corte igual ou maior que 5

Ponto de corte	VP	FN	FP	VN	Sensibilidade	Especificidade	Média
≥ 0	40	0	60	0	100%	0%	50%
≥ 1	40	0	55	5	100%	8,3%	54,2%
≥ 2	39	1	48	12	97,5%	20%	58,8%
≥ 3	39	1	38	22	97,5%	36,7%	67,1%
≥ 4	38	2	29	31	95%	51,7%	73,3%
≥ 5	37	3	15	45	92,5%	75%	83,8%
≥ 6	32	8	10	50	80%	83,3%	81,7%
≥ 7	26	14	6	54	65%	90%	77,5%
≥ 8	19	21	3	57	47,5%	95%	71,3%
≥ 9	12	28	1	59	30%	98,3%	64,2%
= 10	2	38	0	60	5%	100%	52,5%

Para entender o impacto dos diferentes pontos de corte na sensibilidade e na especificidade de um procedimento diagnóstico deve-se reconhecer que em uma população existem dois grupos de pessoas: aquelas que de fato apresentam a condição clínica (grupo clínico) e aquelas não apresentam a condição clínica (grupo saudável). Mais ainda, as pessoas que fazem parte do grupo clínico e do grupo saudável apresentarão diferentes resultados frente ao procedimento diagnóstico, gerando assim duas distribuições distintas. A Figura 1 apresenta duas distribuições fictícias dos resultados dos grupos saudável e clínico em relação a um procedimento diagnóstico qualquer. Notar que embora o grupo saudável tenha o resultado

menor em relação ao grupo clínico neste procedimento diagnóstico, existe a superposição entre essas duas curvas, o que torna difícil a decisão acerca da presença ou da ausência da condição clínica.

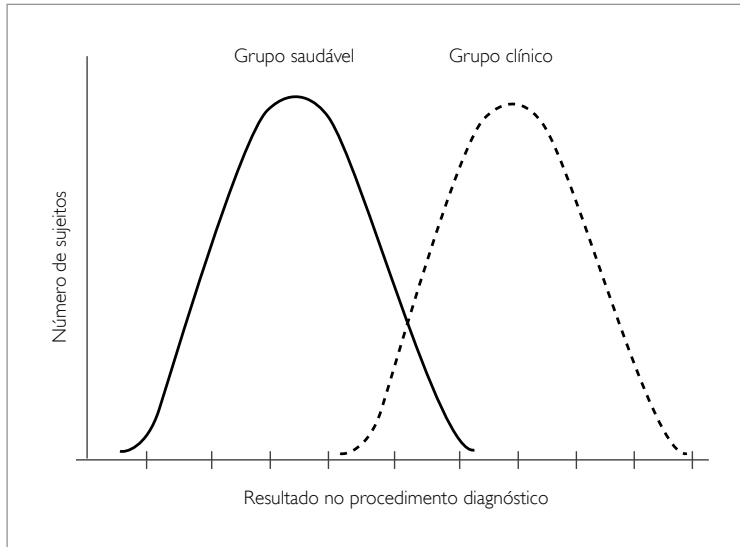


Figura 1 Distribuição de um grupo saudável (ausência da condição clínica) e outro com a presença da condição clínica frente a determinado procedimento diagnóstico.

Pode-se adotar esses princípios ao exemplo cujos resultados encontram-se representados na Tabela 7. Neste caso, há um grupo clínico de pacientes com esclerose múltipla que apresenta comprometimento cognitivo e outro grupo de pessoas que, embora sofram de esclerose múltipla, não apresentam déficit cognitivo, sendo, portanto, o grupo saudável em relação à avaliação cognitiva. A Figura 2 apresenta a distribuição dos resultados desses dois grupos na bateria neuropsicológica breve. Calculada nos resultados apresentados na Tabela 7, a Figura 2 apresenta também as decisões às quais se chegaria caso se adotasse um critério leniente, ou seja, ponto de corte igual ou maior que 2. Neste caso, há alta sensibilidade (taxa de VP igual a 97,5%) e, conseqüentemente, a taxa de FN (100 – sensibilidade) igual a 2,5%. Entretanto, esse ponto de corte gera especificidade muito baixa (taxa de VN igual a 20%) e, conseqüentemente, a alta taxa de FP (100 – especificidade) igual a 80%.

A condição oposta, na qual se adota o ponto de corte extremamente rígido, encontra-se representada na Figura 3. Neste caso, em que o ponto de corte é igual ou maior que 8, há baixa sensibilidade (taxa de VP igual a 47,5%) e, conseqüentemente, a taxa relativamente alta de FN (100 – sensibilidade igual a 52,5%). Entretanto, esse ponto de corte gera especificidade muito alta (taxa de VN igual a 95%) e, conseqüentemente, baixa taxa de FP (100 – especificidade igual a 5%).

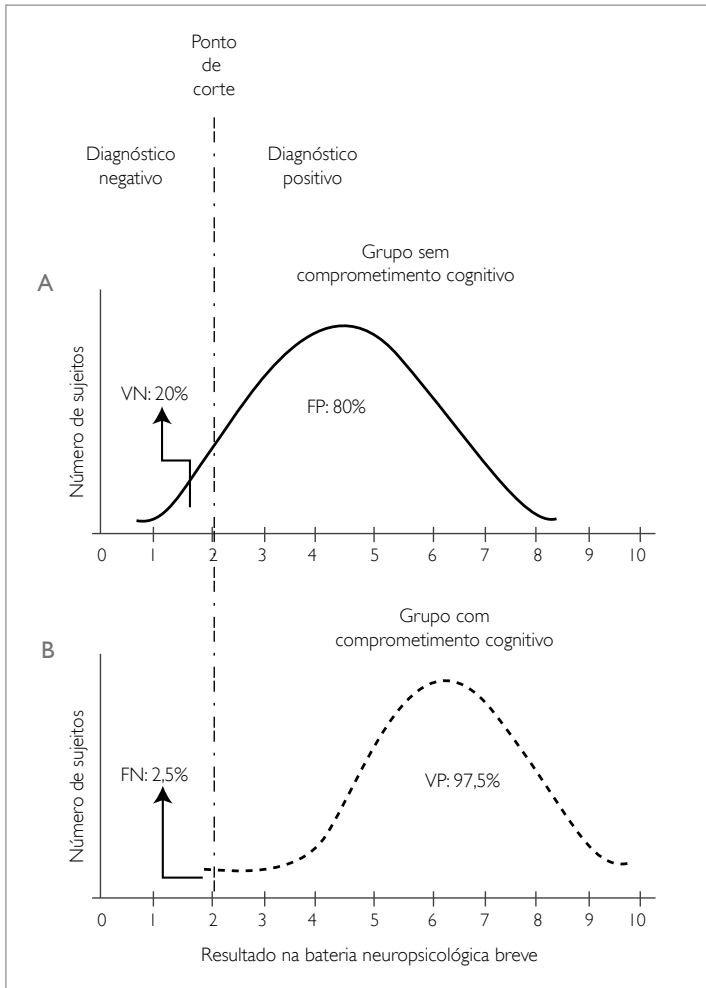


Figura 2 Distribuição dos resultados à bateria neuropsicológica breve em pacientes com esclerose múltipla divididos em dois grupos: um sem comprometimento cognitivo (A) e outro com comprometimento cognitivo (B). A figura ilustra os valores dos valores do VP, FN, FN e VP para ambos os grupos quando se adota o ponto de corte leniente igual o maior que 2 nesta BNB. Neste caso, privilegia-se a sensibilidade do procedimento diagnóstico em detrimento da especificidade.

A Figura 4 representa um ponto de corte intermediário em que se busca maximizar tanto a sensibilidade quanto a especificidade da BNB. Nesse caso, em que o ponto de corte é igual ou maior que 5, há alta sensibilidade (taxa de VP igual a 92,5%) e consequentemente baixa taxa de FN (100 – sensibilidade igual a 7,5%). Esse ponto de corte gera também especificidade bastante alta (taxa de VN igual a 75%) e consequentemente taxa relativamente baixa de FN (100 – especificidade igual a 25%).

A melhor forma de representar o comportamento de cada um desses pontos de corte sobre a sensibilidade e a especificidade de um procedimento diagnóstico é por meio da curva ROC

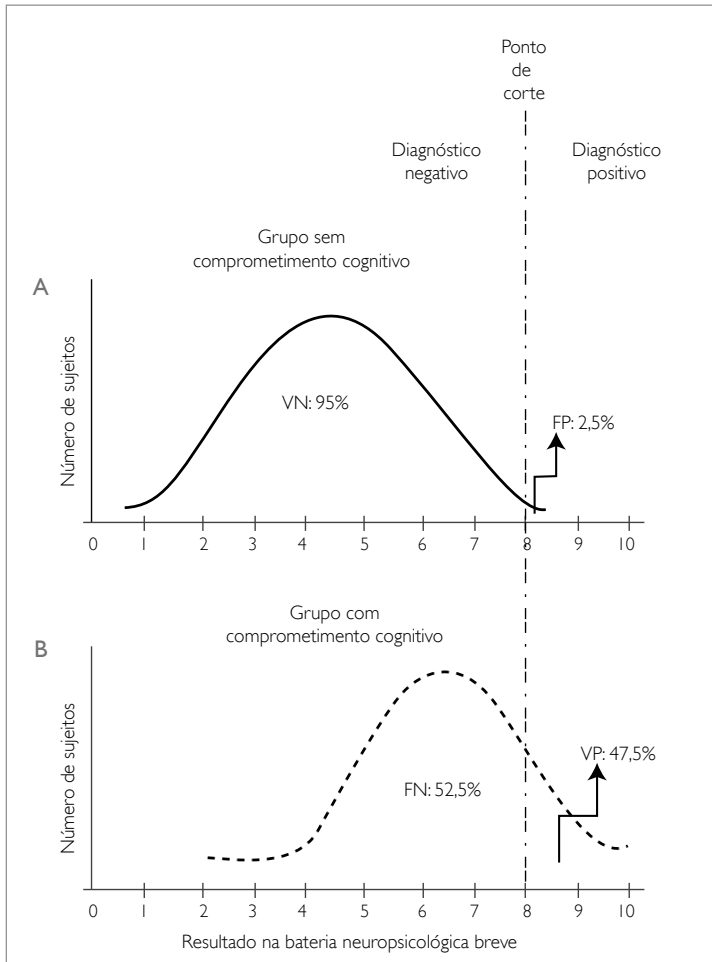


Figura 3 Distribuição dos resultados à BNB em pacientes com esclerose múltipla divididos em dois grupos: um sem comprometimento cognitivo (A) e outro com comprometimento cognitivo (B). A figura ilustra os valores dos valores do VP, FN, FN e VP para ambos os grupos quando se adota um ponto de corte rígido (igual o maior que 8) nesta BNB. Neste caso, privilegia-se a especificidade do procedimento diagnóstico em detrimento da sensibilidade.

(do inglês *receiver operating characteristic*). A curva ROC está calcada na teoria de detecção de sinal desenvolvida durante a II Guerra Mundial para a análise da capacidade de operadores de radar em discriminar a presença ou a ausência de um sinal em relação a um ruído. Daí a nomenclatura “curva característica de operação do receptor”. Na década de 1970, essas técnicas foram incorporadas em diferentes contextos, incluindo a área de diagnósticos clínicos.<sup>4,5</sup>

Para construir uma curva ROC utilizam-se diversos pontos de corte. Cada um desses pontos de corte podem representar no eixo da ordenada (eixo y) o respectivo valor da sensibilidade (taxa de VP), enquanto no eixo da abscissa (eixo x) o respectivo valor da taxa de FP (100 – especificidade). A Figura 5 representa uma típica curva ROC.

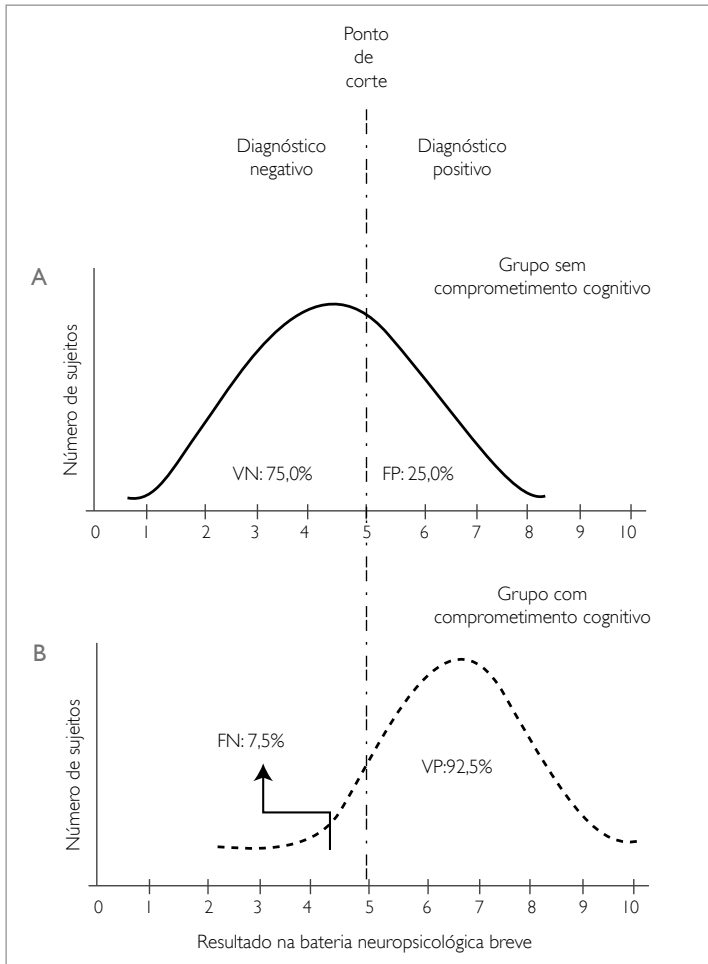


Figura 4 Distribuição dos resultados à BNB em pacientes com esclerose múltipla divididos em dois grupos: um sem comprometimento cognitivo (A) e outro com comprometimento cognitivo (B). A figura ilustra os valores dos valores do VP, FN, VN e FP para ambos os grupos quando se adota um ponto de corte intermediário (igual o maior que 5) nesta BNB. Neste caso, privilegia-se um equilíbrio entre sensibilidade e especificidade desse procedimento diagnóstico.

Conforme se observa na Figura 5, cada ponto da curva ROC representa um par de sensibilidade e especificidade para um ponto de corte específico. Um procedimento diagnóstico perfeito teria a sensibilidade de 100% e especificidade de 100% (ou seja, uma taxa de FP igual a 0%). Dessa forma, quanto mais próxima a curva estiver localizada no canto superior esquerdo, melhor a qualidade ou o desempenho do procedimento diagnóstico em discriminar as pessoas que fazem parte do grupo saudável (ruído) daqueles que fazem parte do grupo clínico (ruído + sinal).

É interessante notar que qualquer ponto sobre a linha diagonal dessa representação gráfica indica que o procedimento diagnóstico não oferece nenhuma informação capaz de

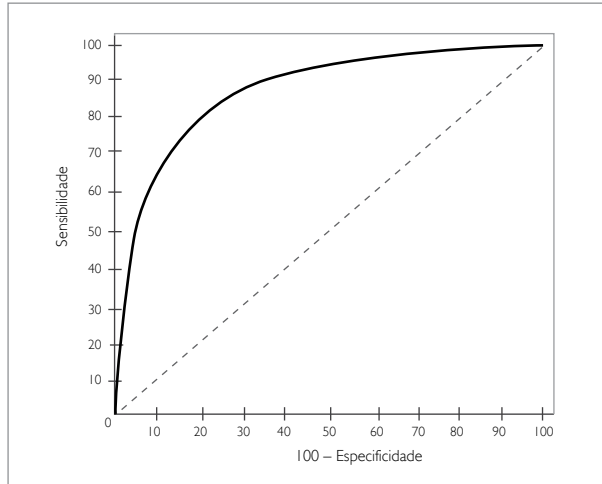


Figura 5 Curva ROC. A taxa de VP (sensibilidade) é representada em função da taxa de FP (100 – especificidade) para diferentes pontos de corte de determinado procedimento diagnóstico.

discriminar esses dois grupos de pessoas. Isso porque os valores da taxa de VP e do FP são sempre iguais para os diferentes pontos de corte, fornecendo assim resultados aleatórios para qualquer pessoa que seja testada. Seria como se se pedisse para que cada paciente com esclerose múltipla jogasse uma moeda para cima. Caso a face da moeda fosse cara, a decisão seria de que o paciente teria ou viria a desenvolver um comprometimento cognitivo (grupo clínico). Caso a face da moeda fosse coroa, o paciente teria um diagnóstico de ausência de comprometimento cognitivo (grupo saudável).

A Figura 6 ilustra a curva ROC dos resultados fictícios apresentados na Tabela 7 (os dados brutos desse exemplo encontram-se no Anexo deste capítulo). Vários programas estatísticos permitem a confecção desse tipo de gráfico. Nesse caso, foi utilizado o programa MedCalc<sup>6</sup> versão 10.2.0.0 ([www.medcalc.org](http://www.medcalc.org)) para analisar esses resultados. Cada ponto dessa curva representa um ponto de corte específico, indo deste critério mais rígido, localizado na porção inferior à esquerda da curva até critérios mais liberais ou lenientes localizados na porção superior à direita da curva.

Critérios mais rígidos caracterizam-se pelo fato de reduzirem bastante os casos de FP, ou seja, privilegiam a especificidade do procedimento diagnóstico, ao custo de deixar de diagnosticar pacientes que de fato apresentam a condição clínica (VP). Ou seja, abrem mão da sensibilidade do procedimento diagnóstico. Este é o caso do ponto A da Figura 6. Critérios mais liberais apresentam características exatamente opostas. Privilegiam a capacidade do procedimento em detectar a presença da condição clínica, ou seja, maximiza-se a taxa de VP (i.e., sensibilidade) ao custo de uma alta taxa de FP. Ou seja, abre-se mão da especificidade desse procedimento. Este é o caso do ponto B da Figura 6.

Embora possam existir vários motivos para adotar critérios mais rígidos ou lenientes, por via de regra adota-se um ponto de corte que permita que o procedimento diagnóstico

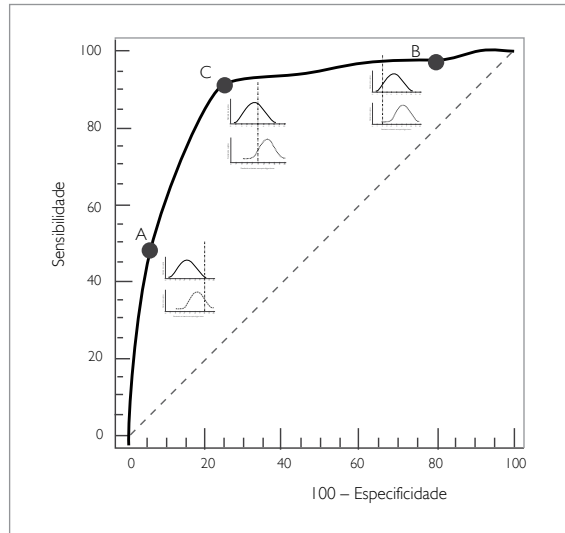


Figura 6 Curva ROC dos resultados fictícios apresentados na Tabela 7. O ponto A representa o ponto de corte muito rígido conforme representado na Figura 3. O ponto B representa o ponto de corte leniente, conforme representado na Figura 2. O ponto C indica o ponto de corte intermediário conforme representado na Figura 4. Este é o ponto de corte que leva ao melhor desempenho da BNB, uma vez que é o que mais se aproxima do canto superior esquerdo da curva.

apresente a melhor capacidade em discriminar os grupos de pessoas que apresentem (grupo clínico) ou não apresentem (grupo saudável) o comprometimento cognitivo. Este é o caso do ponto C da Figura 6, que maximiza tanto a sensibilidade quanto a especificidade da bateria neuropsicológica breve.

#### COMPARAÇÃO ENTRE DIFERENTES PROCEDIMENTOS DIAGNÓSTICOS: A ÁREA SOB A CURVA ROC

Uma mesma condição clínica pode apresentar diferentes procedimentos diagnósticos. Consequentemente é importante avaliar a capacidade que cada um desses procedimentos tem em discriminar as pessoas do grupo saudável do grupo clínico. A Figura 7 apresenta a distribuição desses dois grupos em relação a quatro procedimentos diagnósticos distintos, representadas nas porções A, B, C e D dessa figura. Conforme se observa, quanto mais afastadas estiverem as distribuições do grupo saudável do grupo clínico, maior a capacidade do procedimento diagnóstico em discriminar esses dois grupos.

Pode-se estimar a capacidade que cada procedimento diagnóstico tem em discriminar o grupo clínico do grupo saudável por meio de um índice de discriminabilidade ( $d$ ), cujo valor reflete a diferença entre as médias dos dois grupos dividida pelo desvio-padrão do grupo saudável.<sup>7</sup> Assim, há a subtração entre a média do grupo clínico da média do grupo saudável. O resultado dessa subtração deve ser dividido pelo desvio-padrão do grupo



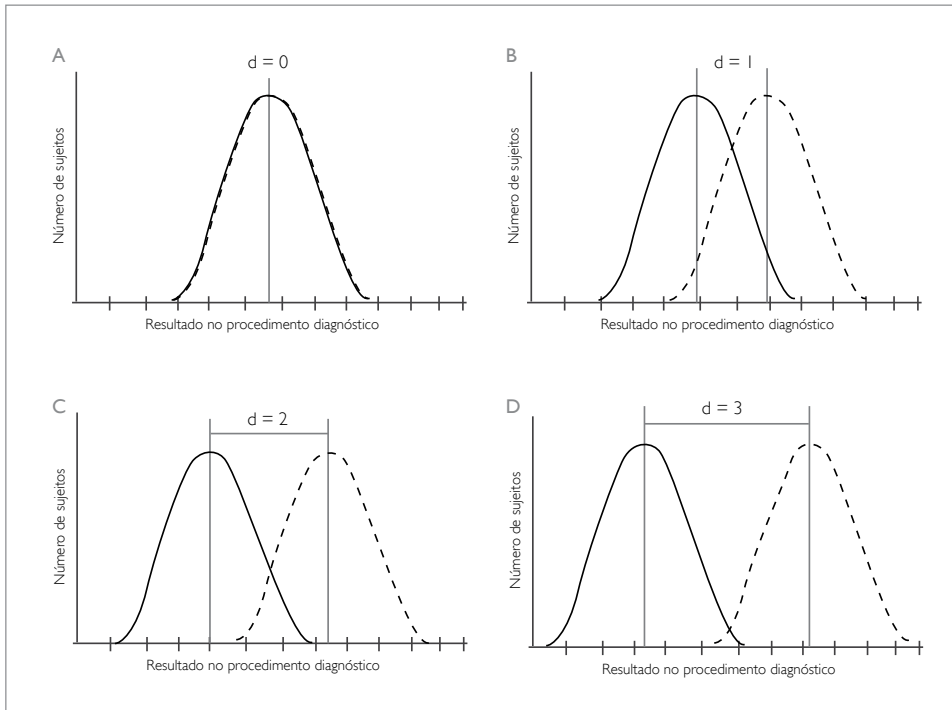


Figura 7 Distribuição dos grupos saudável e clínico em quatro procedimentos diagnósticos diferentes (A, B, C e D) e seus respectivos índices de discriminabilidade ( $d$ ). Quanto mais afastadas estiverem as distribuições do grupo saudável do grupo clínico e maior o índice de discriminabilidade do procedimento diagnóstico, maior a capacidade desse procedimento em discriminar esses dois grupos.

clínico. Quanto maior o resultado dessa divisão, melhor o desempenho do procedimento diagnóstico.

A Figura 7 apresenta alguns exemplos desses índices de discriminabilidade para cada um dos quatro procedimentos diagnósticos. Na Figura 7A pode-se observar uma condição na qual o procedimento diagnóstico gera resultados aleatórios para ambos os grupos, como no exemplo em que o paciente simplesmente joga uma moeda para cima para decidir se ele apresenta ou não um comprometimento cognitivo. Como não existe nenhuma informação nesse procedimento que permita discriminar o grupo saudável do grupo clínico, observa-se uma superposição entre as duas distribuições e um índice de discriminabilidade ( $d$ ) igual a 0. Conforme as distribuições vão se afastando, o desempenho do procedimento diagnóstico vai melhorando e o índice de discriminabilidade vai aumentando.

O desempenho de cada um dos quatro procedimentos diagnósticos pode ser também avaliado por meio da curva ROC. A Figura 8 apresenta quatro curvas ROC dos mesmos quatro procedimentos diagnósticos cujas distribuições encontram-se representadas na Figura 7. O procedimento diagnóstico cuja curva ROC encontra-se na linha diagonal ( $d = 0$ ) não apresenta nenhuma capacidade de discriminar pessoas do grupo saudável do gru-

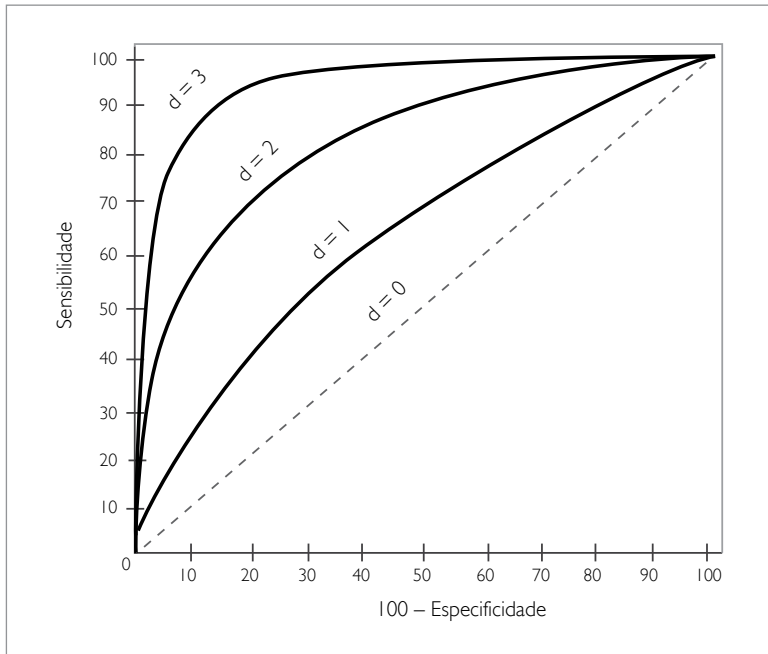


Figura 8 Curvas ROC para os quatro procedimentos diagnósticos apresentados na Figura 7 e seus respectivos índices de discriminabilidade ( $d$ ). Quanto mais próxima a curva estiver do canto superior esquerdo, maior o índice de discriminabilidade e melhor o desempenho do procedimento diagnóstico.

po clínico. Conforme as curvas vão se flexionando para o canto superior esquerdo, maior o índice de discriminabilidade do procedimento diagnóstico e, conseqüentemente, maior a capacidade desse procedimento em discriminar esses dois grupos.

Embora visualmente seja possível constatar as diferenças no desempenho de cada um desses quatro procedimentos diagnósticos, a curva ROC permite quantificar o desempenho de cada um desses procedimentos, calculando-se a respectiva área que ocupa abaixo de sua curva. A área sob a curva ROC representa um dos índices mais usados para avaliar o desempenho de um procedimento diagnóstico, permitindo que ele seja comparado com outros procedimentos diagnósticos ou mesmo em relação a uma condição aleatória na qual não exista nenhuma informação acerca da condição clínica que está sendo investigada.

Quanto maior a área sob a curva ROC, melhor o desempenho do procedimento diagnóstico. A área sob a curva ROC de um procedimento diagnóstico perfeito (ou seja, não há nenhuma superposição na distribuição dos resultados desse procedimento diagnóstico entre o grupo saudável e o grupo clínico) tem valor igual a 1, correspondendo a toda a área da curva, uma vez que a sua taxa de VP (sensibilidade) é igual a 100% e a taxa de FP (100 - especificidade) é igual a 0%. Por outro lado, a área sob a curva do procedimento diagnóstico que fornece resultados positivos ou negativos aleatórios (ou seja, existe uma superposição completa na distribuição dos resultados desse procedimento diagnóstico entre o grupo sau-

dável e o grupo clínico) tem valor igual a 0,5, correspondendo à metade da área, ou seja, a curva é uma diagonal que divide a área pela metade.

O valor da área sob a curva pode ser utilizado para realizar inferências estatísticas acerca do desempenho de um procedimento diagnóstico, testando sua significância estatística em relação a outros procedimentos ou a uma situação aleatória.<sup>8</sup> Dessa forma, pode-se derivar um valor de  $p$  para toda a área sob a curva ROC associada a um procedimento diagnóstico. Esse valor testa a hipótese nula de que o procedimento diagnóstico não seria capaz de distinguir as pessoas entre o grupo saudável e o clínico. Ou seja, os resultados desse procedimento clínico não são aleatórios. Quando existem dois ou mais procedimentos diagnósticos, pode-se testar se existe uma diferença significativa entre esses dois procedimentos, no sentido de que um deles discrimina melhor o grupo saudável do grupo clínico em relação ao outro procedimento diagnóstico. Tradicionalmente, adota-se um nível de significância estatística quando  $p \leq 0,05$ .

Pode-se ilustrar essa técnica estatística com o exemplo da bateria neuropsicológica breve que busca detectar comprometimento cognitivo em pacientes que sofrem de esclerose múltipla. Nesse caso, cada paciente, além de completar a bateria neuropsicológica breve, também é avaliado com o minixame do estado mental, um dos instrumentos mais utilizados para a avaliação de comprometimento cognitivo. O minixame do estado mental (MEEM) gera resultados de números inteiros que podem variar do mínimo de 0 pontos, o qual indica o maior grau de comprometimento cognitivo, até o total máximo de 30 pontos, que corresponde à melhor capacidade cognitiva. Portanto, o MEEM apresenta a direção oposta à BNB, uma vez que quanto maior o valor desse resultado pior o desempenho cognitivo do paciente.

Os resultados fictícios da aplicação desses dois instrumentos em 100 pacientes com esclerose múltipla são apresentados no anexo deste capítulo. Com base nesses resultados é possível construir a curva ROC para cada um dos procedimentos diagnósticos. A Figura 9 apresenta essas duas curvas. Conforme se pode observar nessa figura, a BNB parece ser um procedimento diagnóstico superior ao MEEM, uma vez que sua curva ROC está mais próxima do canto superior esquerdo e possui, aparentemente, uma área maior sob a sua curva ROC em comparação com a curva ROC do MEEM.

Para testar essa hipótese, é possível calcular a área sob a curva ROC de cada um desses procedimentos diagnósticos e compará-los entre si, assim como com uma curva diagonal (área sob a curva ROC igual a 0,5) para verificar se existem diferenças significativas (i.e.,  $p \leq 0,05$ ). Vários pacotes estatísticos podem realizar esses cálculos. Nesse caso, foi utilizada a combinação do MedCalc versão 10.2.0.0 ([www.medcalc.org](http://www.medcalc.org)) e o programa SPSS versão 17.0 (<http://www.spss.com.br>).

A Tabela 8 apresenta a área sob a curva ROC dos resultados dos 100 pacientes com esclerose múltipla frente ao MEEM e à BNB. Ambas as áreas são significativamente maiores do que uma área igual a 0,5 (área sob a curva ROC diagonal), uma vez que o valor de  $p$  em ambos os procedimentos diagnósticos foi menor que 0,05 (MEEM:  $p = 0,01$ ; BNB:  $p <$

0,001). A Tabela 8 indica também que a BNB teve melhor desempenho que o MEEM, uma vez que a comparação entre as áreas sob a curva ROC desses dois procedimentos diagnósticos foi significativamente diferente ( $p < 0,001$ ). A área sob a curva ROC da BNB é maior em 0,235 em relação ao MEEM ( $0,888 - 0,654 = 0,235$ ).

Tabela 8 Área sob a curva ROC do MEEM e da BNB, assim como os valores de p para cada uma dessas áreas. A última linha da tabela apresenta o valor de p para a comparação entre os dois procedimentos diagnósticos

Procedimento diagnóstico	Área sob a curva ROC	Valor de p
MEEM	0,654	. = 0,01
BNB	0,888	< 000,1
MEEM versus BNB	–	< 000,1

Assim, calcular a área sob a curva ROC de um procedimento diagnóstico tem grande valor no processo de desenvolvimento e avaliação de novos procedimentos diagnósticos. Neste caso, ambos os procedimentos (MEEM e BNB) agregaram informações para uma decisão clínica acerca da condição cognitiva em pacientes que sofrem de esclerose múltipla. Mais ainda, a técnica da área sob a curva ROC permitiu também concluir, neste exemplo fictício, que a BNB representa um procedimento diagnóstico superior em relação ao MEEM para o propósito de avaliação da condição cognitiva entre pacientes que sofrem de esclerose múltipla.

## CONSIDERAÇÕES FINAIS

Decisões clínicas arrogadas estão amparadas em procedimentos diagnósticos que permitam inferir a real existência ou não de uma condição clínica de forma clara e precoce. Para que se possa atingir tais objetivos, pesquisas que desenvolvam novos procedimentos diagnósticos são fundamentais. A teoria de detecção de sinal deriva uma série de procedimentos estatísticos que permitem avaliar diferentes aspectos de um procedimento diagnóstico. Entre eles destacam-se sensibilidade, especificidade, valores preditivos positivo e negativo. Finalmente, a curva ROC permite estabelecer o melhor ponto de corte de um novo procedimento diagnóstico, assim como comparar com outros procedimentos diagnósticos. Essas ferramentas estatísticas permitem clarificar as limitações inerentes a qualquer procedimento diagnóstico e consequentemente balizar decisões de profissionais que atuam em ambiente clínico.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. Paes RA, Alvarenga RM, Vasconcelos CC, Negreiros MA, Landeira-Fernandez J. Neuropsicología de la esclerosis múltiple primaria progresiva. Rev Neurologia. 2009;49:343-8.

2. Negreiros MA, Mattos P, Landeira-Fernandez J, Paes RA, Alvarenga RP. A brief neuropsychological screening test battery for cognitive dysfunction in brazilian multiple sclerosis patients. *Brain Inj.* 2008;22:419-26.
3. Loong TW. Understanding sensitivity and specificity with the right side of the brain. *BMJ.* 2003;327(7417):716-9.
4. Swets JA. Measuring the accuracy of diagnostic systems. *Science.* 1988;240(4857):1285-93.
5. Swets JA, Dawes RM, Monahan J. Better decisions through science. *Sci Am.* 2000;283(4):82-7.
6. MedCalc (MedCalc Software, Inc, Mariakerke, Belgium) Disponível em: [medcalc.org/?gclid=CKSZno2woakCFYjt7Qod7k96vQ](http://medcalc.org/?gclid=CKSZno2woakCFYjt7Qod7k96vQ) (acesso 22 set 2011).
7. Simpson AJ, Fitter MJ. What is the best index of detectability? *Psychol Bull.* 1973;80:481-8.
8. Liu H, Li G, Cumberland WG, Wu T. Testing statistical significance of the area under a receiving operating characteristics curve for repeated measures design with bootstrapping. *J Data Science.* 2005;3:257.

## ANEXO

Dados fictícios utilizados nos exemplos das Tabelas 7 e 8 e das Figuras 2, 3, 4, 6 e 9

Sujeito	Diag	BNB	MEEM	Sujeito	Diag	BNB	MEEM
1	1	9	24	51	0	4	28
2	0	5	28	52	1	5	29
3	0	7	27	53	1	9	27
4	0	4	29	54	1	7	27
5	0	5	28	55	1	7	25
6	0	0	29	56	0	4	25
7	1	9	29	57	0	0	30
8	1	8	25	58	1	6	24
9	0	3	28	59	0	6	29
10	0	1	29	60	1	5	30
11	1	1	29	61	0	3	28
12	0	0	29	62	0	2	28
13	0	4	28	63	1	10	20
14	0	1	30	64	1	8	29
15	0	4	27	65	0	3	29
16	0	2	30	66	0	8	27
17	0	4	30	67	1	9	23
18	0	4	29	68	1	9	27
19	1	8	28	69	0	2	25
20	1	9	29	70	0	5	29
21	0	6	25	71	0	8	29
22	0	5	29	72	1	7	27
23	0	0	29	73	0	1	28
24	1	10	23	74	1	7	29
25	0	7	23	75	0	2	30
26	1	9	27	76	0	6	30
27	0	4	29	77	0	2	29

(continua)

Dados fictícios utilizados nos exemplos das Tabelas 7 e 8 e das Figuras 2, 3, 4, 6 e 9 (continuação)

Sujeito	Diag	BNB	MEEM	Sujeito	Diag	BNB	MEEM
28	0	4	27	78	1	6	24
29	1	6	29	79	1	6	27
30	0	1	29	80	0	4	28
31	1	8	27	81	1	6	30
32	0	4	29	82	0	4	26
33	1	7	20	83	0	5	29
34	1	9	25	84	0	3	30
35	1	9	24	85	1	8	29
36	0	3	29	86	1	8	27
37	0	6	27	87	0	9	30
38	1	7	26	88	0	3	29
39	0	7	28	89	0	3	29
40	0	3	29	90	1	6	27
41	1	5	30	91	1	8	29
42	1	4	30	92	0	1	27
43	1	5	27	93	0	0	29
44	0	2	29	94	0	1	27
45	0	2	28	95	1	9	30
46	0	2	29	96	1	7	23
47	1	3	30	97	1	5	28
48	0	1	30	98	0	4	28
49	0	4	29	99	0	2	28
50	0	2	26	100	0	3	29

Diag: diagnóstico acerca da presença ou ausência de comprometimento cognitivo avaliado com extensa bateria neuropsicológica ao longo de várias sessões (padrão-ouro); BNB: bateria neuropsicológica breve; MEEM: minixame do estado mental.