NASCER E CRESCER
revista do hospital de crianças maria pia
ano 2010, vol XIX, n.º 4

Avaliação Crítica e Implementação Prática de Estudos Sobre a Validade de Testes Diagnósticos – Parte II

Luís Filipe Azevedo[1], Altamiro da Costa Pereira[1]

RESUMO

A Medicina Baseada na Evidência (MBE) é genericamente definida como a aplicação consciente, explícita e criteriosa da melhor evidência científica disponível na tomada de decisões sobre o cuidado individual dos doentes. Nesta série de Pediatria Baseada na Evidência têm vindo a ser abordados os aspectos conceptuais, metodológicos e operacionais relativos à prática da MBE. Neste artigo é apresentada a segunda parte do exemplo prático sobre a avaliação crítica e aplicação prática de estudos sobre a validade de testes diagnósticos iniciado no artigo anterior desta série. Nesta segunda parte é dada continuidade à discussão do cenário clínico previamente apresentado e em que existia a necessidade de encontrar e avaliar a evidência sobre a validade e utilidade da procalcitonina sérica na distinção entre pneumonia de etiologia bacteriana e vírica em crianças. No artigo anterior foram discutidos métodos de pesquisa da evidência e foi sugerida uma metodologia sistemática para a avaliação crítica de estudos sobre a validade de testes diagnósticos, incluindo três fases: (1) avaliação da qualidade metodológica do estudo; (2) avaliação da importância científica e prática dos seus resultados e (3) avaliação da aplicabilidade prática dos mesmos. Depois de no artigo anterior terem sido abordadas as questões relativas à primeira destas fases, no presente artigo são abordados os métodos, conceitos e critérios necessários para a avaliação da importância científica e prática dos resultados (características operacionais dos testes diagnósticos e aplicação da informação diagnóstica no processo de tomada de decisão clínica numa perspectiva bayesiana) e da aplicabilidade prática dos mesmos.

Palavras-chave: Medicina baseada na evidência, diagnóstico, testes diagnósticos, especificidade, sensibilidade, valores preditivos, razões de verosimilhança, curvas ROC, procalcitonina, pneumonia, criança.

Nascer e Crescer 2010; 19(4): 265-277

INTRODUÇÃO

A Pediatria Baseada na Evidência foi genericamente definida, no primeiro artigo desta série[1], como a aplicação consciente, explícita e criteriosa da melhor evidência científica disponível na tomada de decisões sobre o cuidado individual dos doentes pediátrico[2-4]. No segundo e terceiro artigos da série[5,6] foram desenvolvidos em maior detalhe os aspectos relacionados com a formulação de questões clínicas, a eficaz pesquisa da evidência científica e a avaliação crítica da mesma. No quarto e quinto artigos[7,8] iniciou-se a apresentação de um conjunto de cenários práticos que exemplificam a aplicação dos métodos e competências abordadas, tendo sido o primeiro exemplo dedicado à avaliação crítica e implementação prática de revisões sistemáticas e estudos de meta-análise e o segundo dedicado aos ensaios clínicos aleatorizados.

No sexto artigo da série[9], foi dado início à apresentação e discussão de um cenário prático que exemplificava a pesquisa, avaliação crítica e implementação prática de estudos sobre a validade de testes diagnósticos. Neste sétimo artigo irá ser concluída a discussão do cenário iniciado no número anterior da série, focando especificamente as questões relacionadas com os métodos, conceitos e critérios necessários para a avaliação da importância científica e prática dos resultados de estudos sobre a validade de testes diagnósticos (características operacionais dos testes diagnósticos e aplicação da informação diagnóstica no processo de tomada de decisão clínica numa perspectiva bayesiana) e para a avaliação da aplicabilidade na prática clínica dos mesmos.

No número anterior desta série[9] iniciamos a apresentação do caso da Maria, de 5 anos de idade, que foi trazida ao serviço de urgência de pediatria com sinais e sintomas sugestivos de pneumonia. O clínico responsável pela Maria – um interno de pediatria – após a recolha da história clínica e a realização do exame físico, procedeu ao pedido de vários exames complementares de diagnóstico. Ao fazer estes pedidos tentou lembrar-se das razões porque considerava necessários cada um dos exames que solicitava, tendo-se tornado claro na sua cabeça que uma das principais questões a que tentava responder era a da etiologia, vírica ou bacteriana, da pneumonia eventualmente existente, dado que isto teria consequências fundamentais na decisão sobre o plano terapêutico. Em particular, relativamente à questão típica e fulcral da necessidade de prescrição de antibióticos. Neste exercício, acabou por lembrar-se de ter ouvido falar recentemente sobre a procalcitonina e a sua potencial utilidade na distinção entre infecções víricas e bacterianas, e não dominando ainda o assunto, decidiu procurar a evidência existente sobre a utilidade diagnóstica

[1] Serviço de Bioestatística e Informática Médica da Faculdade de Medicina da Universidade do Porto e Centro de Investigação em Tecnologias e Sistemas de Informação em Saúde – CINTESIS.

da procalcitonina na discriminação entre pneumonias de etiologia bacteriana e vírica na criança. Após a adequada formulação da questão e de uma pesquisa bibliográfica dirigida à mesma, utilizando a base de dados *Medline*, encontrou uma lista com algumas dezenas de artigos, sendo vários deles altamente relevantes. Logo no início desta lista destacava-se o artigo de Don et al., publicado em 2007, no *Scandinavian Journal of Infectious Diseases* e intitulado "*Efficacy of serum procalcitonin in evaluating severity of community-acquired pneumonia in childhood*"[10]. Após uma rápida leitura de alguns dos resumos encontrados, tornou-se evidente que este era um dos artigos mais recentes e que respondia de forma mais específica à questão. O passo seguinte foi aceder à versão integral do artigo e iniciar a sua análise crítica.

(A) Avaliação da qualidade metodológica do estudo

Ao longo do último artigo desta série[9] foi possível verificar que a qualidade dos estudos sobre validade de testes diagnósticos está dependente do controlo de um vasto conjunto de fontes de erros sistemáticos e aleatórios que devem ser conhecidas, prevenidas e detectadas. Foi definido um conjunto de 14 critérios básicos para a avaliação sistemática da qualidade metodológica deste tipo de estudos (ver Tabela 1), sendo a sua aplicação discutida e ilustrada, utilizando como exemplo prático o artigo de Don et al.[10] sobre a validade da procalcitonina na discriminação entre pneumonias de etiologia bacteriana e vírica na criança. Verificou-se que este artigo, apesar de globalmente adequado quanto à sua qualidade metodológica, tinha três problemas importantes: (1) a utilização de um teste padrão de referência que, sendo genericamente adequado, não conseguia identificar a etiologia específica da pneumonia em 34% dos doentes, com o potencial de enviesamento associado e a consequente redução da amostra a analisar; (2) o estudo tinha um tamanho de amostra limitado, com consequentes estimativas de validade do teste diagnóstico de precisão muito limitada (devido aos potenciais erros aleatórios associados às pequenas amostras) e, por último, (3) os autores não reportavam explicitamente a precisão das suas estimativas (através da apresentação de intervalos de confiança ou erros padrões para as estimativas das características operacionais do teste diagnóstico), não dando oportunidade aos leitores de avaliar criticamente este aspecto.

No presente artigo será discutido o desfecho do cenário clínico apresentado e, à luz da avaliação da importância científica e prática dos resultados do estudo e da aplicabilidade à prática clínica dos mesmos (ver Tabela 1), dando continuidade à avaliação crítica do artigo de Don et al.[10], será discutida a relevância e adequação da utilização da procalcitonina na discriminação entre pneumonias de etiologia vírica e bacteriana na criança.

(B) Avaliação da importância científica e prática dos resultados

O diagnóstico poderá ser genericamente definido como um processo probabilístico e iterativo de decisão que, fazendo o melhor uso possível de toda a informação disponível a cada momento, visa classificar o indivíduo numa determinada entidade nosológica à qual estão associados um determinado plano terapêutico e um determinado prognóstico. O diagnóstico implica, antes de mais, ter a capacidade de discriminar entre indivíduos com e sem a entidade nosológica em análise (por simplicidade, fala-se habitualmente em doentes e não doentes). Neste contexto, definem-se como testes diagnósticos quaisquer dados, medições ou classificações que permitam discriminar entre populações com e sem a entidade nosológica em consideração. Nesta definição genérica incluem-se quer os habituais exames complementares de diagnóstico (imagiológicos, hematológicos, bioquímicos, microbiológicos, imunológicos, genéticos, etc.), quer os dados provenientes da história clínica, antecedentes pessoais e familiares e observações do exame físico.

A utilização adequada e consciente dos testes diagnósticos exige o conhecimento e compreensão das propriedades fundamentais que os caracterizam: (1) reprodutibilidade; (2) validade; (3) efeitos da sua utilização nas decisões clínicas e variáveis de resultado clínico relevantes e (4) aceitabilidade, riscos, custos e custo-efectividade associados. Sendo várias as suas características ou propriedades relevantes, é muito comum colocar uma ênfase especial na validade, entre outras razões, porque é nesta propriedade que assenta a aplicação dos resultados do teste no processo de tomada de decisão clínica, tal como ao longo deste artigo iremos demonstrar. A validade de um teste diagnóstico pode ser definida como a sua capacidade de medir ou classificar aquilo que realmente pretende medir ou classificar. Assim, a validade mede a capacidade do teste diagnóstico classificar adequadamente o indivíduo como doente ou não doente, através da comparação do resultado do teste em avaliação com aquilo que é considerado, num determinado momento no tempo, como sendo o resultado mais próximo da verdade, ou seja, o resultado de um teste diagnóstico padrão de referência (gold standard), aceite consensualmente para a situação em apreço (infelizmente, em não raras situações, não teremos disponível um *gold standard* consensualmente aceite).

A avaliação da importância científica e prática dos resultados dos estudos sobre a validade de testes diagnósticos (ver Tabela 1) passa pela adequada interpretação e contextualização das medidas de validade do teste e subsequente verificação, em última análise, da sua real capacidade de discriminação entre indivíduos com e sem a entidade nosológica em consideração. Genericamente, a avaliação da importância dos resultados deverá considerar dois aspectos essenciais[4,11]:

São apresentadas medidas adequadas à quantificação da validade do teste diagnóstico em avaliação? O teste diagnóstico em avaliação demonstra uma importante capacidade de discriminação?

O valor de qualquer teste diagnóstico está directamente relacionado com a informação diagnóstica adicional que este traz, estando esta directamente relacionada com a sua validade e a subsequente modificação da probabilidade do diagnóstico em consideração trazida pelo seu resultado. Um teste diagnóstico só deve ser utilizado quando o seu resultado

Tabela 1: Critérios de avaliação de estudos sobre a validade de testes diagnósticos relativamente à: (A) qualidade metodológica, (B) importância dos resultados e (C) aplicabilidade prática dos resultados.

(A) Avaliação da qualidade metodológica do estudo:
1. O espectro de doença nos participantes do estudo é representativo do da população que irá, potencialmente, ser submetida ao teste no contexto clínico real de utilização do mesmo
2. Os critérios de selecção dos participantes, o contexto, o local e o período em que o estudo foi realizado são claramente descritos?
3. Os métodos de amostragem dos participantes são claramente descritos? Foi usado um método de amostragem consecutiva seguindo os critérios de selecção estabelecidos?
4. O estudo e a recolha dos dados foram planeados antes da aplicação do teste em avaliação e do teste padrão (estudo prospectivo) ou depois da realização destes (estudo retrospectivo)?
5. O teste padrão de referência (*gold standard*) utilizado é adequado à situação clínica e consegue discriminar adequadamente entre os indivíduos com e sem a entidade nosológica em consideração?
6. O espaço temporal entre a aplicação do teste padrão e do teste diagnóstico em avaliação foi suficientemente curto de modo a poder assumir-se que não existiram alterações relevantes do estado do indivíduo entre estes dois momentos?
7. Todos os participantes, ou uma amostra aleatória dos participantes, foram submetidos ao teste padrão de referência?
8. Todos os participantes receberam o mesmo teste padrão de referência independentemente do resultado do teste diagnóstico em avaliação?
9. O teste padrão de referência é independente do teste diagnóstico em avaliação (i.e. o teste em avaliação não é parte integrante do teste padrão de referência)?
10. A aplicação e execução do teste padrão de referência e do teste em avaliação são descritas em suficiente detalhe que permita a sua replicação?
11. O resultado do teste em avaliação foi obtido e interpretado sem o conhecimento prévio do resultado do teste padrão (ocultação)? O resultado do teste padrão foi obtido e interpretado sem o conhecimento prévio do resultado do teste em avaliação (ocultação)?
12. Quando o resultado do teste em avaliação foi obtido e interpretado, a informação clínica disponível era semelhante àquela que está habitualmente disponível no contexto clínico real de utilização do mesmo?
13. Os resultados dos testes classificados como ambíguos, intermédios, não interpretáveis ou omissos são descritos e os métodos e procedimentos usados no seu tratamento explicados?
14. Os participantes que abandonaram o estudo ou que tiveram um seguimento incompleto são descritos, as razões de perda ou abandono descritas e os métodos e procedimentos usados no seu tratamento explicados?
(B) Avaliação da importância científica e prática dos resultados
15. São apresentadas medidas adequadas à quantificação da validade do teste diagnóstico em avaliação? O teste diagnóstico em avaliação demonstra uma importante capacidade de discriminação?
16. É apresentada a precisão das estimativas das medidas de validade do teste diagnóstico em avaliação?
(C) Avaliação da aplicabilidade prática dos resultados
17. Serão os resultados do estudo aplicáveis ao meu doente, tendo em conta as suas características específicas? Serão os resultados do estudo generalizáveis para o meu contexto específico?
18. O teste diagnóstico em avaliação está disponível e é aplicável, válido e reprodutível no contexto onde me insiro?
19. Estará o doente disposto a ser submetido ao teste diagnóstico? Existem riscos ou eventos adversos eventualmente associados à sua aplicação?
20. Será possível encontrar estimativas clinicamente relevantes das probabilidades pré-teste, de forma a optimizar a utilização da informação diagnóstica proveniente do teste?
21. Serão os resultados do teste diagnóstico úteis para a resolução do problema do meu doente?
22. Quais são as opiniões, valores e expectativas do meu doente relativamente aos objectivos ou resultados clínicos esperados e poderá a realização do teste diagnóstico contribuir para a concretização dos mesmos?

pode modificar a actuação clínica que se segue (atente-se, no entanto, que muitas vezes o pedido de um teste poderá estar relacionado não só com o diagnóstico em consideração, mas com outros objectivos clínicos relevantes, por exemplo, avaliar a gravidade/evolução da doença, avaliar o estado geral do indivíduo, estabelecer um perfil inicial ou estado basal da doença). Se o resultado de um teste não vai ter consequências, a sua utilização é clínica e eticamente discutível. É fundamental, antes de solicitar um teste diagnóstico, perceber claramente o problema de decisão em causa e perceber quando é que, realmente, a informação diagnóstica adicional é relevante nesse contexto. Assim, quanto à evidência científica sobre a validade de testes, é fundamental conhecer as medidas específicas de validade dos testes diagnósticos, verificar se estas são adequadamente reportadas nos estudos sobre este assunto, entender os princípios da aplicação da informação diagnóstica proveniente dos testes e entender o método de actualização de probabilidades de doença após o conhecimento dos resultados dos testes.

Sensibilidade, especificidade e valores preditivos

As medidas de validade de um teste diagnóstico surgem da comparação dos resultados deste com os resultados de um teste padrão de referência. Por uma questão de facilidade, iremos considerar na discussão que se segue a simplificação do resultado do teste diagnóstico classificando-o unicamente como positivo ou negativo. Esta é uma simplificação, já que em grande parte dos casos o resultado do teste diagnóstico será representado em escalas categóricas ordinais ou numéricas, discretas ou contínuas, e não através da simples dicotomia positivo/negativo. O resultado de um teste diagnóstico é geralmente dicotomizado em positivo/negativo através da escolha de pontos de corte (*cut-off points*) adequados na escala do teste. Associadas a pontos de corte diferentes estarão medidas de validade ou características operacionais do teste diferentes, sendo necessário verificar a variação das mesmas em função dos pontos de corte defi-

nidos e saber escolher os pontos de corte mais adequados para o teste, em função dos objectivos do mesmo (ver adiante a secção sobre curvas ROC e área sob a curva ROC). Classicamente, no entanto, e assumindo a dicotomização simplificadora positivo/negativo, são definidas as seguintes medidas de validade ou características operacionais de um teste diagnóstico (Ver Tabela 2)[12,13]:

Sensibilidade: definida como a probabilidade condicional de ter um teste positivo dada a existência de facto de doença e estimada pela proporção de indivíduos que têm um teste positivo de entre os doentes (condicionando o cálculo da proporção só aos doentes).

$$\text{Sensibilidade (Se)} = P[T+ \mid D]$$

$$\hat{Se} = \frac{a}{a+c}$$

Um teste muito sensível é aquele que consegue "apanhar a maior parte dos doentes", ainda que, por esse motivo, possa falsamente identificar alguns não doentes como doentes (falsos positivos). Assim, e porque nestes testes a quase totalidade dos indivíduos com um resultado negativo serão de facto não doentes, um teste muito sensível é mais útil quando é negativo (porque exclui com grande certeza o diagnóstico em consideração). Na literatura anglo-saxónica encontra-se, por vezes, o acrónimo *SnNout* para expressar esta ideia – "*When a test has a high Sensitivity, a Negative result rules out the diagnosis*".

Especificidade: definida como a probabilidade condicional de ter um teste negativo dada a ausência de doença e estimada pela proporção de indivíduos que têm um teste negativo de entre os não doentes.

$$\text{Especificidade (Es)} = P[T- \mid \bar{D}]$$

$$\hat{Es} = \frac{d}{b+d}$$

Um teste muito específico é aquele que consegue "apanhar com grande certeza aqueles que são realmente doentes", ainda que, por esse motivo, possa falsamente identificar alguns doentes como não doentes (falsos negativos). Assim, e porque nestes testes a quase totalidade dos indivíduos com testes positivos serão de facto doentes, um teste muito específico é mais útil quando é positivo (porque identifica com grande certeza o diagnóstico em consideração). Na literatura anglo-saxónica encontra-se, por vezes, o acrónimo *SpPin* para expressar esta ideia – "*When a test has a high Specificity, a Positive result rules in the diagnosis*".

Valor preditivo positivo: definido como a probabilidade condicional de ter a doença dado que o resultado do teste foi positivo e estimada pela proporção de indivíduos que têm a doença de entre os que têm um teste positivo.

$$\text{Valor Preditivo Positivo (VPP)} = P[D \mid T+]$$

$$\hat{VPP} = \frac{a}{a+b}$$

Uma vez solicitado o teste e recebido o seu resultado a informação clinicamente mais relevante, e que para o doente é mais importante, é a da probabilidade de ter a doença dado o resultado observado do teste. Neste contexto a probabilidade de ter de facto a doença dado que foi observado um resultado positivo do teste corresponde ao valor preditivo positivo.

Valor preditivo negativo: definido como a probabilidade condicional de não ter a doença dado que o resultado do teste foi negativo e estimada pela proporção de indivíduos que não têm a doença de entre os que têm um teste negativo.

$$\text{Valor Preditivo Negativo (VPN)} = P[\bar{D} \mid T-]$$

$$\hat{VPN} = \frac{d}{c+d}$$

Num registo semelhante ao do valor preditivo positivo, o valor preditivo negativo dá-nos a noção, clinicamente mais relevante, da probabilidade de um indivíduo não estar realmente doente quando o teste foi negativo.

Razões de verosimilhança (likelihood ratios - LR)

Outras medidas de validade que podem ser calculadas para um teste diagnóstico, que têm particular importância pela sua utilidade no cálculo das probabilidades de doença actualizadas após o conhecimento do resultado do teste, são as razões de verosimilhança (*likelihood ratio – LR*).

Define-se a razão de verosimilhança (LR) para cada resultado possível de um teste diagnóstico, como sendo a razão entre a probabilidade desse resultado na população doente e a probabilidade desse resultado na população não doente, isto é:

$$Likelihood\ ratio\ (T) = \frac{P[T|D]}{P[T|\bar{D}]}$$

Assim, para um resultado positivo do teste, tem-se que o LR+ corresponde a:

$$Likelihood\ ratio\ (T+) = LR+ =$$

$$= LR+ = \frac{P[T+|D]}{P[T+|\bar{D}]} = \frac{Se}{1-Es}$$

Tabela 2 – Representação habitual dos resultados do teste diagnóstico cuja validade se pretende avaliar em função dos resultados do teste padrão de referência, que determinam o verdadeiro estado de doente e não doente para cada participante.

		Doença		
		Presente	Ausente	
Teste	Positivo	a	b	a+b
	Negativo	c	d	c+d
		a+c	b+d	

E para um resultado negativo do teste, tem-se que o LR- corresponde a:

$$Likelihood\ ratio\ (T-) = LR- =$$

$$= LR- = \frac{P[T-|D]}{P[T-|\overline{D}]} = \frac{1-Se}{Es}$$

O LR+ diz-nos quantas vezes mais provável é encontrar um resultado positivo do teste nos doentes em comparação com os não doentes. Portanto, este será tanto maior quanto mais sugestivo da presença de doença for um resultado positivo do teste. Valores de LR+ entre 2 e 5 condicionam pequenas alterações da probabilidade pré-teste; valores entre 5 e 10 condicionam alterações moderadas e valores maiores que 10 condicionam grandes alterações[14-16].

O LR- diz-nos quantas vezes menos provável é encontrar um resultado negativo do teste nos doentes em comparação com os não doentes. Portanto, este será tanto menor quanto mais sugestivo da ausência da doença for um resultado negativo do teste. Valores do LR- entre 0,5 e 0,2 condicionam pequenas alterações da probabilidade pré-teste; valores entre 0,2 e 0,1 condicionam alterações moderadas e valores menores que 0,1 condicionam grandes alterações[14-16].

Actualização da probabilidade de doença e teorema de Bayes

Verifica-se que a utilização de um determinado teste diagnóstico serve precisamente para alterar uma determinada probabilidade *a priori* da doença ($P[D]$ – probabilidade pré-teste), de forma a obter uma estimativa dessa probabilidade que tenha em conta a informação dada pelo resultado do teste diagnóstico ($P[D|T+]$ ou $P[D|T-]$ – probabilidades pós-teste – genericamente $P[D|T]$). A estimativa de $P[D|T]$ deverá ser obtida a partir da probabilidade pré-teste da doença $P[D]$ e das características operacionais básicas do teste (sensibilidade $P[T+]$ e especificidade $P[T-|\overline{D}]$) que são, à partida, os únicos dados conhecidos.

A metodologia adequada para a actualização da probabilidade de doença dado o resultado de um teste diagnóstico passa pela aplicação do teorema de Bayes. Este teorema resulta da própria definição de probabilidade condicional e permite obter uma expressão para a probabilidade de um acontecimento *a posteriori* (por exemplo, $P[D|T]$) com base na sua probabilidade *a priori* e na probabilidade condicional invertida (por exemplo, $P[D]$ e $P[T|D]$, probabilidade pré-teste e sensibilidade e/ou especificidade do teste)[12, 13].

Sejam $P[T+]$, $P[T-]$, $P[D]$ e $P[\overline{D}]$ as probabilidades do resultado de um teste diagnóstico ser positivo e negativo, e as probabilidades de presença e ausência da doença (dada pelo resultado do teste padrão de referência). Sejam ainda $P[D|T+]$, $P[D|T-]$, $P[T+|D]$ e $P[T-|\overline{D}]$, respectivamente, a probabilidade condicional de ter doença dado que se teve um resultado positivo do teste (probabilidade pós-teste positivo), a probabilidade condicional de ter doença dado que se teve um resultado negativo (probabilidade pós-teste negativo), a probabilidade condicional de ter um resultado positivo dado que se é doente (sensibilidade) e a probabilidade condicional de ter um resultado negativo dado que se é não doente (especificidade). Tem-se, por definição, que uma probabilidade condicional é dada pela razão entre a probabilidade conjunta dos eventos (intersecção) e a probabilidade do evento condicionante. Ou seja, por exemplo, a probabilidade condicional de estar doente dado que se tem um teste positivo ($P[D|T+]$) é igual à razão entre a probabilidade conjunta de estar doente e ser positivo ($P[D\cap T+]$) e a probabilidade de ter um teste positivo ($P[T+]$). E, da mesma forma, a probabilidade condicional de ter um teste positivo dado que se é doente ($P[T+|D]$) é igual à razão entre a probabilidade conjunta de estar doente e ser positivo ($P[D\cap T+]$) e a probabilidade de ser doente ($P[D]$). Assim:

$$P[T+|D] = \frac{P[D\cap T+]}{P[D]}$$

e

$$P[D|T+] = \frac{P[D\cap T+]}{P[T+]}$$

Tendo em conta as definições e notação atrás apresentadas o teorema de Bayes poderá ser deduzido da seguinte forma:

Partindo das expressões anteriores verificamos que:

$$P[D\cap T+] = P[T+|D]\times P[D]$$

e

$$P[D\cap T+] = P[D|T+]\times P[T+]$$

Logo,

$$P[T+|D]\times P[D] = P[D|T+]\times P[T+] \Leftrightarrow$$

$$\Leftrightarrow P[D|T+] = \frac{P[T+|D]\times P[D]}{P[T+]} \Leftrightarrow$$

$$\Leftrightarrow P[D|T+] = \frac{P[T+|D]\times P[D]}{P[T+\cap D]+P[T+\cap\overline{D}]} \Leftrightarrow$$

$$\Leftrightarrow P[D|T+] =$$

$$= \frac{P[T+|D]\times P[D]}{P[T+|D]\times P[D]+P[T+|\overline{D}]\times P[\overline{D}]}$$

Assim, para um resultado positivo ou negativo do teste, tem-se respectivamente:

$$P[D|T+] =$$

$$= \frac{P[T+|D]\times P[D]}{P[T+|D]\times P[D]+P[T+|\overline{D}]\times P[\overline{D}]}$$

$$P[D|T-] =$$

$$= \frac{P[T-|D]\times P[D]}{P[T-|D]\times P[D]+P[T-|\overline{D}]\times P[\overline{D}]}$$

Onde, tendo em conta as definições atrás apresentadas, é possível verificar que:

$$P[T-|D] = 1-P[T+|D] =$$

$$= 1-Sensibilidade\ (Se)$$

$$P[T+|\overline{D}] = 1-P[T-|\overline{D}] =$$

$$= 1-Especificidade\ (Es)$$

E simplificando, através do uso dos conceitos de sensibilidade e especificidade, conclui-se finalmente que[12, 13]

$$P[D|T+] = \frac{Se \times P[D]}{Se \times P[D] + (1-Es) \times (1-P[D])}$$

$$P[D|T-] = \frac{(1\text{-}Se) \times P[D]}{(1\text{-}Se) \times P[D] + Es \times (1-P[D])}$$

Considere-se o seguinte exemplo a título ilustrativo. Um homem tem 55 anos, apresenta hemoptises e perda ponderal nos últimos 6 meses e é fumador há mais de 20 anos (cerca de 2 maços por dia). Baseado na história clínica e exame físico, o clínico coloca como primeira hipótese diagnóstica uma neoplasia pulmonar, e estima, com base nos dados clínicos, que a probabilidade pré-teste de neoplasia pulmonar é, neste caso, de 0,4 ou 40%. O clínico solicita uma radiografia pulmonar como parte do estudo deste doente e esta revela a presença de uma massa no lobo superior direito. Apesar de haver alguma heterogeneidade na evidência existente, será justo afirmar que a radiografia pulmonar tem sensibilidade de cerca de 60% e uma especificidade de cerca de 85% no diagnóstico de neoplasias pulmonares[17-19].

Neste cenário, como deverá o clínico interpretar o achado na radiografia pulmonar? Qual será a utilidade deste teste? De que forma poderá o resultado positivo do teste (presença de massa pulmonar) alterar a probabilidade de doença estimada à partida pelo clínico?

Tendo em conta a probabilidade pré-teste de doença e a sensibilidade e especificidade do teste, a aplicação do teorema de Bayes permite responder a estas questões:

$$P[D|T+] = \frac{\mathrm{Se} \times P[D]}{\mathrm{Se} \times P[D] + (1-\mathrm{Es}) \times (1-P[D])}$$

$$P[D|T+] = \frac{0,6 \times 0,4}{(0,6 \times 0,4) + (0,15 \times 0,6)}$$

$$P[D|T+] = \frac{0,24}{0,24+0,09} = \frac{0,24}{0,33} = 0,73$$

$$P[D|T+] = 0,73$$

Antes da informação dada pelo teste, o clínico estimava em 40% a probabilidade de presença de neoplasia pulmonar (probabilidade pré-teste) e após a verificação do resultado do teste essa probabilidade é actualizada para 73% (probabilidade pós-teste positivo). O resultado positivo do teste (presença de massa pulmonar) altera moderadamente a probabilidade da doença.

Se pelo contrário, a radiografia pulmonar não fosse sugestiva da presença de neoplasia pulmonar, a probabilidade de doença, dado o resultado negativo do teste, seria calculada da seguinte forma:

$$P[D|T-] = \frac{(1\text{-}\mathrm{Se}) \times P[D]}{(1\text{-}\mathrm{Se}) \times P[D] + \mathrm{Es} \times (1-P[D])}$$

$$P[D|T-] = \frac{0,4 \times 0,4}{(0,4 \times 0,4) + (0,85 \times 0,6)}$$

$$P[D|T-] = \frac{0,16}{0,16+0,51} = \frac{0,16}{0,67} = 0,24$$

$$P[D|T-] = 0,24$$

Neste caso, a probabilidade de doença seria diminuída relativamente à estimativa inicial feita pelo clínico antes de pedir o teste. O clínico passaria de uma probabilidade de doença de 40% para uma probabilidade de 24%.

Este exemplo torna claro o facto, intuitivamente evidente, de que a radiografia pulmonar é, neste contexto, um teste imperfeito, que quando negativo não exclui convincentemente a doença e quando positivo necessita sempre de um teste mais específico como complemento (tipicamente a tomografia computarizada).

Odds de doença, ***likelihood ratios*** (LR) e teorema de Bayes

Uma forma alternativa e muito mais apelativa de representar o teorema de Bayes, é através da aplicação dos conceitos de *Odds* de doença e das razões de verosimilhança (*likelihood ratios – LR*) do teste diagnóstico[12, 13].

O *Odds* de uma doença é, simplesmente, uma forma alternativa de representar a probabilidade da mesma e é definida como a razão entre a probabilidade da doença e a probabilidade complementar de não existência de doença:

$$Odds[D] = \frac{P[D]}{P[\overline{D}]}$$

e

$$P[D] = \frac{Odds[D]}{1+Odds[D]}$$

O conceito de *likelihood ratio (LR*) foi já definido acima.

Considerem-se as expressões do teorema de Bayes para presença e ausência da doença, dado um determinado resultado do teste (*T*, positivo ou negativo):

$$P[D|T] = \frac{P[T|D] \times P[D]}{P[T|D] \times P[D] + P[T|\overline{D}] \times P[\overline{D}]}$$

$$P[\overline{D}|T] = \frac{P[T|\overline{D}] \times P[\overline{D}]}{P[T|\overline{D}] \times P[\overline{D}] + P[T|D] \times P[D]}$$

Se agora for considerada a razão destas duas expressões obtém-se uma nova igualdade com particular interesse:

$$\frac{P[D|T]}{P[\overline{D}|T]} = \frac{P[T|D] \times P[D]}{P[T|D] \times P[D] + P[T|\overline{D}] \times P[\overline{D}]} \Big/ \frac{P[T|\overline{D}] \times P[\overline{D}]}{P[T|\overline{D}] \times P[\overline{D}] + P[T|D] \times P[D]}$$

E após simplificação, esta poderá ser escrita da seguinte forma mais compacta:

$$\frac{P[D|T]}{P[\overline{D}|T]} = \frac{P[D]}{P[\overline{D}]} \cdot \frac{P[T|D]}{P[T|\overline{D}]}$$

Atendendo às definições de *Odds* e *Likelihood ratio (LR)* atrás apresentadas, tem-se que esta expressão poderá ser escrita da seguinte forma[12, 13]:

$$Odds[D|T] = Odds[D] \times Likelihood\ ratio\ [T]$$

Isto é, o *Odds* de doença após o conhecimento do resultado do teste (*Odds* pós-teste) é igual ao produto do *Odds* de doença antes do teste (Odds pré-teste) e do *likelihood ratio (LR)* para o resultado do teste obtido.

Dada a simplicidade desta última expressão e a disponibilidade, cada vez mais frequente, de valores da razão de verosimilhança (*likelihood ratio – LR*) para os vários resultados possíveis dos testes diagnósticos na literatura, esta forma alternativa do teorema de Bayes é crescentemente utilizada, na actualização da probabilidade da doença após obtenção do resultado do teste diagnóstico. Este é o método mais simples e prático de aplicação do teorema de Bayes.

Tendo em atenção o exemplo atrás apresentado, o *Odds* de doença pós-teste, no caso do teste ter um resultado positivo (radiografia com massa pulmonar), poderia ser calculado da seguinte forma:

$$Odds[D|T+] =$$

$$= Odds[D] \times Likelihood\ ratio[T+]$$

$$Odds[D|T+] =$$

$$= \frac{0,4}{0,6} \cdot \left(\frac{Se}{1-Es} \right) = 0,67 \cdot \frac{0,60}{0,15} = 2,67$$

$$Odds[D|T+] = 2,67$$

Tendo em conta a relação entre *Odds* e probabilidade, tem-se neste caso:

$$Odds[D|T+] =$$

$$= \frac{P[D|T+]}{1-P[D|T+]} \Leftrightarrow P[D|T+] =$$

$$= \frac{Odds[D|T+]}{1+Odds[D|T+]}$$

$$P[D|T+] = \frac{2,67}{1+2,67} = 0,73$$

$$P[D|T+] = 0,73$$

Confirmando, assim, o resultado obtido anteriormente e que apontava para uma probabilidade pós-teste positivo de 73%. O mesmo aconteceria se fosse calculada, com este método, a probabilidade pós-teste negativo.

Curvas ROC e área sob a curva ROC

O resultado de um teste diagnóstico, tipicamente, será representado em escalas categóricas ordinais ou numéricas, discretas ou contínuas, e não através da simples dicotomia positivo/negativo que foi assumida nas secções anteriores deste artigo. O resultado de um teste diagnóstico é geralmente dicotomizado em positivo/negativo através da escolha de pontos de corte (*cut-off points*) assumidos na escala do teste. Associadas a pontos de corte diferentes estarão medidas de validade ou características operacionais do teste diferentes, sendo necessário verificar a variação das mesmas em função dos pontos de corte definidos e saber escolher os pontos de corte mais adequados em função dos objectivos do teste.

Quando consideramos os resultados de um teste diagnóstico numa escala ordinal ou numérica, ao escolher um determinado ponto de corte (que determina os resultados ditos positivos e negativos), geralmente, a sensibilidade e a especificidade são características difíceis de conciliar, isto é, é complicado aumentar a sensibilidade e a especificidade ao mesmo tempo. As curvas ROC (*receiver operator characteristic curve*) são uma forma de representar a relação, normalmente antagónica, entre a sensibilidade e a especificidade, ao longo de um conjunto de valores de corte na escala do teste a avaliar (*cut-off points*).

Para construir uma curva ROC traça-se uma curva que une os vários pontos num diagrama representando a sensibilidade em função da proporção de falsos positivos (1- Especificidade) para um conjunto de valores de corte (*cut-off points*) distintos (ver exemplos na figura 1).

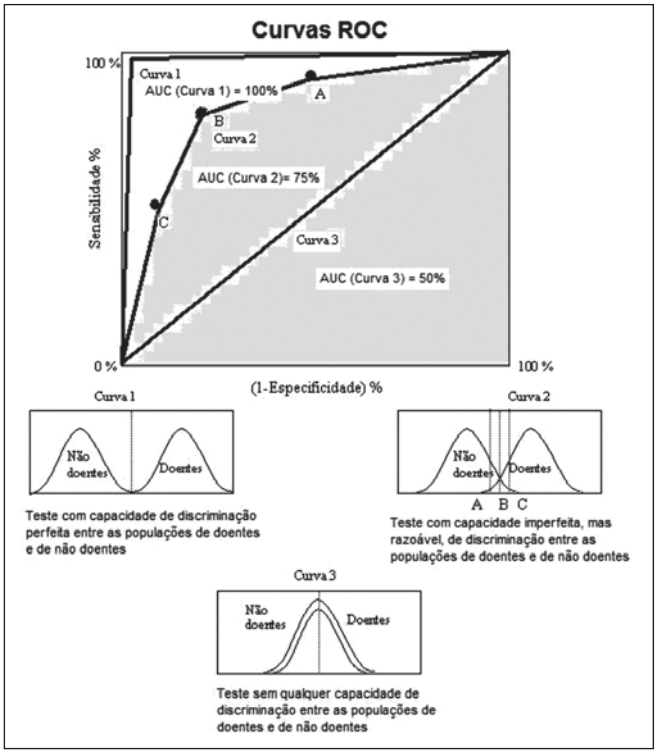

Figura 1 – Representação das curvas ROC e respectivas áreas sob a curva ROC (*area under the curve – AUC*) para três testes com capacidade de discriminação distintas. Teste com capacidade de discriminação perfeita entre doentes e não doentes (curva 1), teste com discriminação imperfeita (curva 2) e teste sem qualquer capacidade de discriminação (curva 3)

O valor escolhido como ponto de corte (*cut-off point*) vai influenciar as características operacionais do teste, como exemplificado na figura 1 (curva 2). No exemplo desta figura, quanto maior é o ponto de corte maior é a especificidade do teste, mas menor é a sensibilidade (ponto C da curva 2); e quanto menor o ponto de corte maior é a sensibilidade, mas menor é a especificidade (ponto A da curva 2). Percebe-se, desta forma, que a intenção com que se utilizará o teste diagnóstico influencia a escolha do melhor ponto de corte, logo, das características do teste. No exemplo da curva 2 da figura 1, se pretendemos um teste muito sensível e menos específico, escolhe-se um ponto de corte menor (ponto A), obtendo-se uma menor proporção de falsos negativos e uma maior proporção de falsos positivos; se pretendemos, pelo contrário, um teste muito específico e menos sensível, escolhe-se um ponto de corte maior (ponto C), obtendo-se uma menor proporção de falsos positivos e uma maior proporção de falsos negativos.

As curvas ROC descrevem a capacidade de discriminação de um determinado teste diagnóstico para um conjunto diverso de valores de corte. Isto permite pôr em evidência os valores de corte para os quais é possível fazer uma optimização da sensibilidade em função da especificidade. O ponto, numa curva ROC, onde se consegue optimizar conjuntamente os valores de sensibilidade e especificidade do teste é aquele que se encontra mais próximo do canto superior esquerdo do diagrama (ver figura 1, ponto B da curva 2).

Para além disto, as curvas ROC permitem quantificar a exactidão de um teste diagnóstico (medida da capacidade global de discriminação de um teste diagnóstico), já que, esta é proporcional à área sob a curva ROC de um determinado teste (também designada de estatística C ou *area under the curve – AUC*), isto é, será tanto maior quanto mais a curva se aproxima do canto superior esquerdo do diagrama. Sabendo isto, a curva será útil na percepção da capacidade global de discriminação de um determinado teste e na comparação entre testes diagnósticos, tendo um teste uma exactidão tanto maior quanto maior for a sua área sob a curva ROC (ver figura 1, neste caso, o teste correspondente à curva 1 tem melhor discriminação global – exactidão – do que os testes das curvas 2 e 3, e da mesma forma, o teste da curva 2 é melhor que o da curva 3). A área sob a curva ROC é tipicamente apresentada na forma de uma proporção entre 0,5 e 1 ou de percentagem entre 50 e 100%, sendo 50% o valor correspondente a uma curva ROC de um teste que não tem qualquer capacidade de discriminação (curva 3 na figura 1) e 100% o valor correspondente a um teste que discrimina perfeitamente entre doentes e não doentes (curva 1 na figura 1). Habitualmente os testes diagnósticos têm valores de áreas sob a curva ROC entre estes valores extremos, correspondendo a situações de discriminação imperfeita (curva 2 da figura 1).

Um outro método por vezes utilizado para a avaliação da capacidade global de discriminação de um teste diagnóstico, no contexto da análise de curvas ROC, é o chamado índice de Youden (IY)[20]. Este é dado pela seguinte expressão:

$$IY = Max_c\left[Se(c) + Es(c) - 1\right]$$

Isto é, o IY corresponde ao valor máximo da expressão (soma da sensibilidade e da especificidade menos um) quando esta é avaliada para cada possível ponto de corte do teste diagnóstico (c). O teste será tanto melhor na discriminação entre doentes e não doentes quanto maior for o IY. Da mesma forma, o IY é útil na definição do ponto que melhor maximiza simultaneamente sensibilidade e especificidade, uma vez que este corresponde, precisamente, ao ponto de corte (c) que torna máxima a expressão acima[20].

Os resultados apresentados no estudo de Don et al.[10], cuja avaliação crítica tem vindo a ser feita, dão-nos conta de várias medidas de validade do teste em avaliação – procalcitonina – na discriminação entre pneumonias de etiologia bacteriana e vírica na criança. Em particular, é possível verificar na tabela IV e V do artigo as sensibilidades, especificidades, valores preditivos positivos e LR+ para cada um dos valores de corte 0,5 ng/ml, 1,0 ng/ml e 2,0 ng/ml na discriminação entre pneumonia de etiologia bacteriana pneumocócica e etiologia vírica e para a discriminação entre pneumonias bacterianas por agentes atípicos e pneumonias víricas, respectivamente. Fazendo a reanálise dos dados apresentados pelos autores na tabela III, de forma a avaliar a validade da procalcitonina na discriminação entre pneumonias de etiologia bacteriana (pneumocócicas mais atípicas) e as de etiologia vírica é possível calcular sensibilidades de 74%, 61% e 54% e especificidades de 35%, 61% e 74%, respectivamente para os valores de corte 0,5 ng/ml, 1,0 ng/ml e 2,0 ng/ml. Os LR+ correspondentes são, respectivamente, 1,14 1,55 e 2,05.

As estimativas dos LR apresentadas são extremamente úteis na prática e permitem-nos concluir que, por exemplo, num cenário em que a probabilidade pré-teste de pneumonia de etiologia bacteriana seja, à semelhança do encontrado neste estudo, de 65% (43/66) a probabilidade pós-teste será no máximo modificada, no caso de um valor da procalcitonina acima de 2 ng/ml (positivo), para 79%. Os cálculos que permitem chegar a esta conclusão são semelhantes aos previamente apresentados e explicados:

$$Odds\left[D|T+\right] =$$

$$= Odds\left[D\right] \times Likelihood\ ratio\left[T+\right]$$

$$Odds\left[D|T+\right] = \frac{0,65}{0,35} \cdot 2,05 = 3,81$$

$$Odds\left[D|T+\right] = 3,81$$

$$P\left[D \mid T+\right] = \frac{Odds\left[D|T+\right]}{1 + Odds\left[D|T+\right]}$$

$$P\left[D \mid T+\right] = \frac{3,81}{1+3,81} = 0,79$$

$$P\left[D \mid T+\right] = 0,79$$

Os autores do artigo apresentam-nos também na secção de resultados o valor da área sob a curva ROC para a procalcitonina na discriminação en-

tre pneumonias de etiologia bacteriana (pneumocócicas mais atípicas) e as de etiologia vírica (Figura 2 do artigo[10]). A área sob a curva ROC encontrada foi de 0,629 ou 62,9%.

Quer os resultados das estimativas de sensibilidade e especificidade, quer os valores dos LR e das subsequentes modificações da probabilidade do diagnóstico em consideração, quer, por último, o valor reportado da área sob a curva ROC da procalcitonina leva-nos a concluir que a capacidade de discriminação da procalcitonina entre pneumonias de etiologia bacteriana e vírica na criança é pobre, não tendo sido detectado neste estudo nenhum valor de corte da procalcitonina com boa capacidade de discriminação.

É apresentada a precisão das estimativas das medidas de validade do teste diagnóstico em avaliação?

As medidas de validade apresentadas num estudo desta natureza são sempre estimativas de um verdadeiro valor desse parâmetro na população em causa, resultantes da análise da amostra de participantes no estudo (subconjunto finito da população, geralmente muito menor que esta). Naturalmente, uma estimativa de validade, por exemplo uma sensibilidade do teste, resultante do estudo de uma amostra de 500 participantes, neste caso considerando só os indivíduos com a doença uma vez que a sensibilidade é a probabilidade condicional de ter um teste positivo dado que se é doente, será mais precisa (menos sensível aos erros aleatórios relacionados com o processo de amostragem) que uma estimativa proveniente de um estudo com 50 participantes. A forma mais comum de reportar a precisão da estimativa de uma medida de validade é através da apresentação do seu intervalo de confiança, geralmente de 95% (IC 95%) das estimativas. Quanto mais estreitos forem os intervalos de confiança em torno da estimativa pontual do parâmetro mais precisa é a estimação do mesmo. Assim, é fundamental perceber qual a qualidade e precisão das estimativas das medidas de validade apresentadas, através da avaliação dos seus IC 95%, já que disso dependerá a interpretação e potencial aplicação prática que lhes podemos dar. Deverá ser dada atenção ao facto de, por exemplo, tipicamente a sensibilidade de um teste diagnóstico ter estimativas menos precisas do que as da especificidade, dado que a sensibilidade é estimada dentro do subgrupo dos indivíduos doentes (geralmente em menor número) e a especificidade é estimada no subgrupo dos não doentes, dado que esta segunda corresponde à probabilidade condicional de ter um teste negativo dado que se é não doente.

Relativamente a este ponto, no estudo de Don et al.[10] existe uma importante falha e limitação, pois não são reportadas, para nenhuma das medidas de validade, as correspondentes medidas da precisão das estimativas, por exemplo, na forma dos seus intervalos de confiança. Para entender a importância desta limitação atente-se que, por exemplo, a estimativa da especificidade da procalcitonina na discriminação entre pneumonias de etiologia bacteriana (pneumocócicas mais atípicas) e as de etiologia vírica, para o valor de corte de 1,0 ng/ml, é de 61%, com IC 95% [41% – 81%]. Se isto tivesse sido devidamente reportado era possível perceber que a estimativa pontual apresentada é bastante imprecisa, sendo o valor encontrado compatível com níveis de especificidade tão dispares como 41% e 81%, se o erro aleatório relacionado com o processo de selecção da amostra fosse tido em consideração. Logicamente, esta limitação está relacionada com o pequeno tamanho da amostra de crianças que participaram no estudo.

(C) Avaliação da aplicabilidade prática da evidência

A última questão no processo de avaliação crítica da evidência será a questão da aplicabilidade prática da mesma (ver tabela 1). O objectivo do profissional de saúde será a eventual aplicação da evidência científica aos seus problemas clínicos e aos seus doentes, logo, a avaliação da aplicabilidade prática é uma questão fulcral neste contexto[4, 11, 15, 16]. O objectivo fundamental será o adequado enquadramento da evidência científica, dadas as características específicas do doente, do profissional de saúde e do contexto da sua prática clínica. Os critérios mais úteis para a avaliação da aplicabilidade prática de estudos sobre validade de testes diagnósticos são os seguintes[4, 11, 15, 16]:

Serão os resultados do estudo aplicáveis ao meu doente, tendo em conta as suas características específicas? Serão os resultados do estudo generalizáveis para o meu contexto específico?

A avaliação da aplicabilidade dos resultados de um estudo desta natureza passa, em primeira instância, pela análise crítica dos critérios de selecção (inclusão e exclusão) utilizados. Diferenças sócio-demográficas e clínicas entre populações podem levar a que as estimativas de validade variem bastante. As medidas da validade dos testes diagnósticos poderão ter utilidade clínica limitada se o espectro de doença nos participantes do estudo não é representativo daquele encontrado na população que irá, potencialmente, ser submetida ao teste no contexto clínico real de utilização do mesmo. Os estudos mais úteis são aqueles que incluem participantes que reflectem adequadamente a diversidade encontrada no contexto prático real. A avaliação crítica das critérios de selecção implica a comparação das características dos meu doentes (sócio-demográficas, biológicas e patológicas) com as características dos participantes do estudo, na tentativa de verificar se as primeiras são de tal forma diferentes das segundas que tornem os resultados não aplicáveis ao meu contexto particular.

No estudo de Don et al.[10] verificamos que o contexto, critérios de selecção e métodos de recrutamento dos participantes sugerem que os resultados poderão ser seguramente generalizáveis para crianças de qualquer país europeu vistas no serviço de urgência, com suspeita de pneumonia e sem patologias crónicas associadas.

O teste diagnóstico em avaliação está disponível e é aplicável, válido e reprodutível no contexto onde me insiro?

A avaliação da aplicabilidade dos resultados passa também, obviamente, pela verificação de que o teste em causa está disponível no nosso contexto específico, é aplicável (existem protocolos técnicos adequados implementados e o teste é seguro e economicamente sustentável) e tem reprodutibilidade e valida-

de adequadas (garantias da qualidade da execução do teste no nosso contexto específico, por exemplo, na minha própria instituição de saúde).

Relativamente à determinação sérica da procalcitonina, existem empresas e laboratórios no nosso país que disponibilizam as técnicas adequadas e que têm capacidade técnica adequada na execução deste teste. Um importante senão deste teste são os seus custos, uma vez que, neste momento, são mais de 3 vezes superiores ao de outros testes mais antigos usados nas mesmas circunstâncias e indicações (ex: proteína C reactiva, velocidade de sedimentação, etc.).

Estará o doente disposto a ser submetido ao teste diagnóstico? Existem riscos ou eventos adversos eventualmente associados à sua aplicação?

A avaliação da aplicabilidade do teste diagnóstico passa também pela consideração da invasividade do teste, dos riscos inerentes à sua execução e dos eventos adversos que este poderá potencialmente condicionar. Importa notar que, muitas vezes, estas características não são adequadamente reportadas nos estudos, mas constituem aspectos muito relevantes e que podem ter implicações na adesão ao teste por parte do doente.

No caso da procalcitonina, este ponto específico não é especialmente preocupante, dado que os riscos associados a este teste estão unicamente ligados à necessidade de uma colheita de sangue periférico igual à realizada em qualquer outra situação da rotina de um serviço de saúde. Na criança, logicamente, mesmo esta colheita deverá ser adequadamente ponderada devido à invasividade do procedimento, mas tipicamente é uma manobra segura.

Será possível encontrar estimativas clinicamente relevantes das probabilidades pré-teste, de forma a optimizar a utilização da informação diagnóstica proveniente do teste?

Vimos atrás que a utilidade de um teste diagnóstico está directamente ligada à modificação que o seu resultado pode condicionar na probabilidade da doença em consideração e, subsequentemente, da atitude clínica que daí decorre. Vimos também os métodos utilizados na prática para poder preceder formalmente ao cálculo das probabilidades de doença após o conhecimento do resultado do teste (probabilidade pós-teste) e vimos que para poder aplicar esses métodos seria preciso ter uma estimativa, mais ou menos precisa, da probabilidade pré-teste da doença (estimada pela sua prevalência). Assim sendo, a disponibilidade de estimativas da prevalência da doença em consideração o mais válidas e precisas possível e, idealmente, específicas do meu contexto particular permitirá optimizar a utilização da informação diagnóstica proveniente do teste. Para obter estas estimativas (1) poderão utilizar-se os resultados dos próprios estudos da validade dos testes diagnósticos (que geralmente incluem ou permitem o cálculo deste tipo de medidas); (2) poderão procurar-se estudos específicos sobre a prevalência da doença em causa, idealmente feitos num contexto e com métodos passíveis de permitir a generalização para a nossa própria realidade ou (3) poder-se-á procurar obter, através de uma análise mais ou menos formal, dados no nosso próprio contexto (ex: no nosso hospital) sobre a prevalência da doença em causa (ex: através da consulta de históricos clínicos, bases de dados clínicas ou administrativas ou, idealmente, desenvolvendo um trabalho científico específico para responder a esta questão).

No caso da procalcitonina na discriminação entre pneumonias de etiologia bacteriana e vírica na criança, será necessário ter uma estimativa da frequência das pneumonias de etiologia bacteriana (doença em consideração neste caso) entre as crianças com quadros compatíveis com o diagnóstico de pneumonia. Este valor pode ser obtido, por exemplo, analisando os dados do próprio artigo de Don et al.[10], cuja análise crítica tem vindo a ser feita, e onde é possível verificar que nas crianças com sinais e sintomas compatíveis com pneumonia, e nas quais foi possível determinar a etiologia da mesma, a frequência de pneumonias bacterianas foi de 65% (43/66). Uma outra forma de aplicar devidamente os dados relativos à validade da procalcitonina seria tentar estimar para o nosso contexto particular (no meu hospital ou no meu centro de saúde) a frequência das pneumonias bacterianas entre as crianças vistas com quadros compatíveis com pneumonia.

Serão os resultados do teste diagnóstico úteis para a resolução do problema do meu doente?

A avaliação da aplicabilidade de um estudo sobre validade de um teste diagnóstico implica também perceber até que ponto a aplicação do teste na prática clínica, e a subsequente modificação da probabilidade da doença associada aos seus resultados, tem um efeito significativo na melhoria dos resultados clínicos que temos por objectivo atingir para aquele doente particular (ex: probabilidade e tempo de sobrevivência, capacidade funcional, qualidade de vida, etc.). Obviamente, o ideal seria que um teste fosse só usado se realmente é necessário e tem um efeito relevante nessas variáveis de resultado. Formalmente, este tipo de questões obrigam a pensar nos testes diagnósticos como intervenções, (semelhantes a outras, por exemplo, as terapêuticas) mas de natureza diagnóstica, e exigem tipologias diferentes das discutidas neste artigo para o seu adequado estudo (por exemplo, ensaios clínicos de diagnóstico – *diagnostic randomized clinical trials – D-RCT's*)[21].

Quais são as opiniões, valores e expectativas do meu doente relativamente aos objectivos ou resultados clínicos esperados e poderá a realização do teste diagnóstico contribuir para a concretização dos mesmos?

Finalmente, a decisão sobre a aplicabilidade prática do estudo sobre a validade de um teste diagnóstico deverá ter em conta o equilíbrio entre benefícios e riscos associados ao teste em cada doente particular e incorporar, também, os seus valores e expectativas. Dever-se-á incorporar a valorização individual que o doente faz sobre os potenciais benefícios e riscos associados ao teste e aos seus resultados como factores de ponderação na comparação entre testes alternativos e na decisão de usar ou não cada teste diagnóstico. Para isto, poder-se-ão utilizar métodos mais ou menos quantitativos e mais ou menos explícitos de elicitação e integração das valorizações dos doentes[4].

CONCLUSÃO

O diagnóstico é um dos mais nobres actos da prática médica. A utilização adequada dos testes diagnósticos depende do conhecimento das suas propriedades e características operacionais, em particular da sua validade, e da aplicação de métodos que permitam maximizar a utilização da informação que a partir deles é possível obter. Ao longo do presente e do anterior artigos desta série foi possível verificar que a qualidade dos estudos sobre validade de testes diagnósticos está dependente do controlo de um vasto conjunto de fontes de erros sistemáticos e aleatórios, que devem ser conhecidas, prevenidas e detectadas. A aplicação de uma conjunto de critérios básicos para a avaliação da qualidade metodológica destes estudos, bem como da importância dos seus resultados e da sua aplicabilidade prática foi discutida e ilustrada através de um exemplo prático. Por fim, e no contexto da aplicação prática das medidas de validade dos testes diagnósticos, foram também discutidos os princípios quantitativos que permitem integrar a informação dada pelo resultado de um teste no processo de decisão diagnóstica.

Nestes dois artigos colocou-se o foco na utilização da informação isolada proveniente de um teste diagnóstico, no entanto, convém sublinhar que hoje, na investigação diagnóstica, é cada vez mais frequente a utilização de estratégias e métodos multivariados, que consideram de forma mais adequada a multiplicidade de informação diagnóstica que está disponível em cada situação e a capacidade de discriminação adicional que, neste contexto, cada teste diagnóstico poderá acrescentar[4, 22]. Referimo-nos, em particular, a métodos mais recentes em investigação diagnóstica, de que são exemplo as chamadas regras de decisão clínica (*clinical decision rules*)[4, 22].

O estudo de Don et al.[10] sobre a validade da procalcitonina na discriminação entre pneumonias de etiologia bacteriana e vírica na criança foi apresentado e discutido ao longo dos dois últimos artigos desta série. Verificou-se que o artigo de Don et al.[10], apesar de globalmente adequado quanto à sua qualidade metodológica tinha alguns aspectos que poderiam ser melhorados ou corrigidos. Quanto à importância e à aplicabilidade prática dos resultados deste artigo, foi possível concluir que, à luz da evidência isolada do mesmo, a utilidade diagnóstica da procalcitonina na discriminação entre pneumonias de etiologia vírica e bacteriana na criança é relativamente limitada.

Em conclusão, apesar dos resultados deste artigo de Don et al.[10] e outros estudos[23, 24] indicarem uma utilidade limitada da procalcitonina neste contexto clínico, a verdade é que existem vários outros estudos que chegam a resultados bastante mais positivos[25-28], encontrando, por exemplo, valores de área sob a curva ROC, relativamente à discriminação entre pneumonias de etiologia vírica e bacteriana, de 94%[25], 93%[26] e 76%[28], bastante superiores aos reportados por Don et al.[10]. No entanto, é importante sublinhar que os estudos com resultados menos positivos têm, na generalidade, melhor qualidade metodológica e maior capacidade de generalização.

Em síntese, e tendo em conta na sua globalidade a evidência heterogénea disponível, será justo dizer que, embora não sendo um teste perfeito, a procalcitonina poderá ter alguma utilidade na discriminação entre pneumonias de etiologia bacteriana e vírica na criança. Isto torna-se particularmente verdade se compararmos a utilidade da procalcitonina, neste contexto, com outros testes alternativos (ex: proteína C reactiva, velocidade de sedimentação, contagem de leucócitos ou interleucina 6), alguns deles frequentemente utilizados na prática clínica. Se a procalcitonina não é perfeita, a verdade é que a evidência existente aponta claramente para a sua superioridade relativamente a todas essas outras alternativas[25, 27-30]. Por fim, se a sua utilidade não é muito clara na definição da etiologia da pneumonia, a evidência existente torna claro que: (1) quando são obtidos valores elevados deste marcador (> 20-30 ng/ml) quase sempre são infecções bacterianas e muitas vezes graves; (2) a procalcitonina é um excelente marcador da gravidade da pneumonia e da necessidade de internamento hospitalar e (3) a procalcitonina é um bom marcador de outras patologias bacterianas graves como meningite e sépsis[25-34].

Voltando finalmente ao cenário clínico apresentado, o interno que atendia a Maria, de 5 anos e com sinais e sintomas compatíveis com pneumonia, acaba por decidir, depois da análise do artigo de Don et al.[10], que necessita de despender algum tempo a analisar a restante literatura e a treinar a aplicação do teorema de Bayes no cálculo de probabilidades de doença pós-teste. Depois da pesquisa e leitura que fez fica, no entanto, com a convicção que, por um lado, alguns dos testes que habitualmente pede nestas situações (PCR e contagem de leucócitos) poderão ser menos úteis do que pensava e, por outro lado, embora a procalcitonina possa ter algumas vantagens marginais sobre essas alternativas, só depois de uma ponderação dos custos deste teste poderia considerar, por exemplo, trocar por rotina o seu pedido da PCR por um de procalcitonina. Decidiu então ir no dia seguinte ao laboratório de bioquímica do hospital perguntar quanto é que custava uma análise de procalcitonina em comparação com a sua já conhecida e velha PCR. Quanto à Maria, com a chegada dos resultados dos exames verificou-se que tinha leucocitose de 20000/ml, neutrofilia de 71%, PCR de 35 mg/l e a radiografia torácica com imagem de condensação no lobo inferior direito. Foi instituída antibioterapia em tratamento domiciliar, com vigilância sintomática e recomendada posterior reavaliação no centro de saúde.

Encerramos sublinhando que, na sequência do acima exposto, e em nossa opinião, a alteração e constante evolução dos processos e tecnologias de diagnóstico e terapêutica são um elemento fundamental à melhoria continuada da qualidade do exercício da medicina, sendo a medicina baseada na evidência uma abordagem que permitirá apoiar e informar, da forma mais adequada, esta evolução e a consequente modificação criteriosa e justificada dos processos, atitudes, tecnologias e práticas clínicas instaladas.

CRITICAL APPRAISAL AND PRACTICAL IMPLEMENTATION OF DIAGNOSTIC TESTS ACCURACY STUDIES – PART II

ABSTRACT

Evidence Based Medicine (EBM) is generically defined as the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. This section of Evidence Based Paediatrics has been covering the conceptual, methodological and operational issues related to the practice of EBM in the field of Paediatrics. In the present article we will continue the discussion of the practical example regarding the critical appraisal and practical implementation of diagnostic tests accuracy studies initiated in the previous article of this section. In the present article we will continue the discussion of a clinical scenario where we have searched for and critically appraise scientific evidence about the accuracy of serum procalcitonin in the distinction of bacterial from viral pneumonia in children. In the last article methods for searching the current best evidence were discussed and a systematic approach for the critical appraisal of diagnostic tests accuracy studies was suggested, including: (1) assessment of the methodological quality of the study; (2) assessment of the scientific and practical impact of its results and (3) assessment of their practical applicability. After discussing in the past article the first of this points, in the present article we will be covering methods and concepts concerning the assessment of scientific and practical impact of results (operational characteristics of diagnostic tests and integration of diagnostic information in the clinical decision making process in a bayesian perspective) and their practical applicability.

Keywords: Evidence based medicine, diagnosis, diagnostic tests, specificity, sensitivity, predictive values, likelihood ratios, ROC curves, procalcitonin, pneumonia, children.

Nascer e Crescer 2010; 19(4): 265-277

BIBLIOGRAFIA

1. Azevedo LF, Costa Pereira A: Pediatria Baseada na Evidência. *Nascer e Crescer Revista do Hospital de Crianças Maria Pia* 2007, vol 16(1):29-31.
2. Evidence-Based Medicine Working Group: Evidence-based medicine. A new approach to teaching the practice of medicine. Evidence-Based Medicine Working Group. *Jama* 1992, 268(17):2420-5.
3. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS: Evidence based medicine: what it is and what it isn't. *Bmj* 1996, 312(7023):71-2.
4. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB: *Evidence-based medicine : how to practice and teach EBM* 2nd edition. Edinburgh: Churchill Livingstone; 2000.
5. Azevedo LF, Costa Pereira A: Pediatria Baseada na Evidência - Formulação de Questões e Pesquisa Bibliográfica. *Nascer e Crescer - Revista do Hospital de Crianças Maria Pia* 2007, 16(3):135-40.
6. Azevedo LF, Costa Pereira A: Pediatria Baseada na Evidência - Avaliação Crítica e Implementação Prática da Evidência. *Nascer e Crescer Revista do Hospital de Crianças Maria Pia* 2008, 17(1):30-6.
7. Azevedo LF, Costa Pereira A: Pediatria Baseada na Evidência - Avaliação Crítica e Implementação Prática de Revisões Sistemáticas e Estudos de Meta-Análise. *Nascer e Crescer Revista do Hospital de Crianças Maria Pia* 2008, 17(4):30-8.
8. Azevedo LF, da Costa Pereira A: Pediatria Baseada na Evidência - Avaliação Crítica e Implementação Prática de Ensaios Clínicos Aleatorizados. *Nascer e Crescer - Revista do Hospital de Crianças Maria Pia* 2009, 18(2):106-19.
9. Azevedo LF, Costa Pereira A: Pediatria Baseada na Evidência - Avaliação Crítica e Implementação Prática de Estudos sobre a Validade de Testes Diagnósticos - Parte I. *Nascer e Crescer - Revista do Hospital de Crianças Maria Pia* 2010, 19(2):116-24.
10. Don M, Valent F, Korppi M, Falleti E, De Candia A, Fasoli L, Tenore A, Canciani M: Efficacy of serum procalcitonin in evaluating severity of community-acquired pneumonia in childhood. *Scand J Infect Dis* 2007, 39(2):129-37.
11. Guyatt G: *Users' guides to the medical literature : essentials of evidence-based clinical practice*. 2nd edition. New York McGraw-Hill Medical; 2008.
12. Hunink MGM, Glasziou P, Siegel J, Weeks J, Pliskin J, Elstein A, Weinstein MC: *Decision making in health and medicine : integrating evidence and values*. Cambridge ; New York: Cambridge University Press; 2001.
13. Sox HC, Blatt MA, Higgins MC, Marton KI: *Medical decision making*. Philadelphia: American College of Physicians; 2007.
14. Bhandari M, Guyatt GH: How to appraise a diagnostic test. *World J Surg* 2005, 29(5):561-6.
15. Jaeschke R, Guyatt G, Sackett DL: Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *Jama* 1994, 271(5):389-91.
16. Jaeschke R, Guyatt GH, Sackett DL: Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *Jama* 1994, 271(9):703-7.
17. Gavelli G, Giampalma E: Sensitivity and specificity of chest X-ray screening for lung cancer: review article. *Cancer* 2000, 89(11 Suppl):2453-6.
18. Quekel LG, Goei R, Kessels AG, van Engelshoven JM: Detection of lung cancer on the chest radiograph: impact of previous films, clinical information, double reading, and dual reading. *J Clin Epidemiol* 2001, 54(11):1146-50.
19. Woodring JH: Pitfalls in the radiologic diagnosis of lung cancer. *AJR Am J Roentgenol* 1990, 154(6):1165-75.
20. Fluss R, Faraggi D, Reiser B: Estimation of the Youden Index and its

associated cutoff point. *Biom J* 2005, 47(4):458-72.
21. de Graaff JC, Ubbink DT, Tijssen JG, Legemate DA: The diagnostic randomized clinical trial is the best solution for management issues in critical limb ischemia. *J Clin Epidemiol* 2004, 57(11):1111-8.
22. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS: Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *Jama* 2000, 284(1):79-84.
23. Korppi M, Remes S: Serum procalcitonin in pneumococcal pneumonia in children. *Eur Respir J* 2001, 17(4):623-7.
24. Korppi M, Remes S, Heiskanen-Kosma T: Serum procalcitonin concentrations in bacterial pneumonia in children: a negative result in primary healthcare settings. *Pediatr Pulmonol* 2003, 35(1):56-61.
25. Gendrel D, Raymond J, Coste J, Moulin F, Lorrot M, Guerin S, Ravilly S, Lefevre H, Royer C, Lacombe C *et al*: Comparison of procalcitonin with C-reactive protein, interleukin 6 and interferon-alpha for differentiation of bacterial vs. viral infections. *Pediatr Infect Dis J* 1999, 18(10):875-81.
26. Moulin F, Raymond J, Lorrot M, Marc E, Coste J, Iniguez JL, Kalifa G, Bohuon C, Gendrel D: Procalcitonin in children admitted to hospital with community acquired pneumonia. *Arch Dis Child* 2001, 84(4):332-6.
27. Toikka P, Irjala K, Juven T, Virkki R, Mertsola J, Leinonen M, Ruuskanen O: Serum procalcitonin, C-reactive protein and interleukin-6 for distinguishing bacterial and viral pneumonia in children. *Pediatr Infect Dis J* 2000, 19(7):598-602.
28. Prat C, Dominguez J, Rodrigo C, Gimenez M, Azuara M, Jimenez O, Gali N, Ausina V: Procalcitonin, C-reactive protein and leukocyte count in children with lower respiratory tract infection. *Pediatr Infect Dis J* 2003, 22(11):963-8.
29. Lorrot M, Moulin F, Coste J, Ravilly S, Guerin S, Lebon P, Lacombe C, Raymond J, Bohuon C, Gendrel D: [Procalcitonin in pediatric emergencies: comparison with C-reactive protein, interleukin-6 and interferon alpha in the differentiation between bacterial and viral infections]. *Presse Med* 2000, 29(3):128-34.
30. Simon L, Gauvin F, Amre DK, Saint-Louis P, Lacroix J: Serum procalcitonin and C-reactive protein levels as markers of bacterial infection: a systematic review and meta-analysis. *Clin Infect Dis* 2004, 39(2):206-17.
31. Gervaix A, Pugin J: [Usefulness of procalcitonin in adults and children]. *Rev Med Suisse* 2005, 1(13):872-4, 877.
32. Casado Flores J, Blanco Quiros A: [Procalcitonin. A new marker for bacterial infection]. *An Esp Pediatr* 2001, 54(1):69-73.
33. Herd D: In children under age three does procalcitonin help exclude serious bacterial infection in fever without focus? *Arch Dis Child* 2007, 92(4):362-4.
34. Schutzle H, Forster J, Superti-Furga A, Berner R: Is serum procalcitonin a reliable diagnostic marker in children with acute respiratory tract infections? A retrospective analysis. *Eur J Pediatr* 2009, 168(9):1117-24.

CORRESPONDÊNCIA
lazevedo@med.up.pt.