

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

*Luís Filipe Azevedo<sup>1</sup>, Altamiro da Costa Pereira<sup>1</sup>*

## RESUMO

Na rubrica de Pediatria Baseada na Evidência têm vindo a ser abordados os aspectos conceptuais, metodológicos e operacionais relativos à prática da MBE no âmbito específico da Pediatria. Neste artigo é apresentado um exemplo prático de aplicação dos conceitos, métodos e competências abordados nos artigos anteriores, debruçando-se especificamente sobre a avaliação crítica e aplicação prática de estudos sobre a validade de testes diagnósticos. É apresentado um cenário clínico em que surge a necessidade de encontrar e avaliar a evidência científica existente sobre a validade e utilidade da determinação da procalcitonina na distinção entre pneumonia de etiologia bacteriana e vírica em crianças. São discutidos os métodos de pesquisa da evidência mais actual e apropriada à questão; e é sugerida uma metodologia sistemática para a avaliação crítica de estudos sobre a validade de testes diagnósticos, incluindo três fases distintas: (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. Neste artigo será desenvolvido o processo até à avaliação da qualidade metodológica do estudo, sendo no próximo número desta rubrica abordadas as questões relacionadas com a avaliação quantitativa e qualitativa da importância dos resultados e a avaliação da sua aplicabilidade prática.

**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/(2): 116-24

## INTRODUÇÃO

A Pediatria Baseada na Evidência foi genericamente definida, no primeiro artigo desta série<sup>(1)</sup>, como a aplicação consciente, explícita e criteriosa da melhor evidência científica disponível na tomada de decisões sobre o cuidado individual dos doentes pediátricos<sup>(2-4)</sup>. No segundo e terceiro artigos da série<sup>(5,6)</sup> 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<sup>(7,8)</sup> 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 e sétimo artigos da série, será apresentado um cenário prático no qual é exemplificada a pesquisa, avaliação crítica e implementação prática de estudos sobre a validade de testes diagnósticos. No sexto artigo será apresentado o cenário prático e serão discutidas questões relacionadas com a pesquisa bibliográfica dirigida à questão clínica formulada e com a avaliação crítica de estudos sobre a validade de testes diagnósticos, no respeitante à avaliação

da qualidade metodológica dos mesmos. No próximo número desta rubrica, no sétimo artigo da série, será completada a discussão da avaliação crítica dos estudos sobre a validade de testes diagnósticos, focando as questões relacionadas com a avaliação quantitativa e qualitativa da importância dos resultados e a avaliação da sua aplicabilidade prática.

## CENÁRIO CLÍNICO

Maria, de 5 anos de idade, é trazida pelas sua mãe ao serviço de urgência de pediatria com febre, há cerca de 8 dias, tosse e expectoração abundante e de coloração amarelada, sem sangue. Refere dor tipo pelurítico na face anterior do hemitórax direito, agravada pela tosse. Há cerca de uma semana teve rinorreia e odinofagia que entretanto melhoraram. A mãe sublinha o agravamento da prostração, anorexia e febre nas últimas 48h. Nega queixas urinárias ou abdominais. No exame físico a Maria apresenta-se hipotónica, tem febre de 39,2°C e não apresenta sinais de dificuldade respiratória. À auscultação pulmonar é possível verificar crepitações e roncos dispersos bilateralmente e presença de sopro tubar no lobo inferior direito, associado a broncofonia. Tem abdómen mole e depressível e não tem sinais meníngeos. A Maria não tem outras alterações relevantes no exame físico, é uma criança sem antecedentes patológicos ou familiares relevantes e vive com a mãe, que está desempregada, vive só, tem baixa escolaridade e poucos recursos económicos.

O clínico responsável pela Maria – um interno de pediatria – após a recolha da história clínica e a realização do exame físico senta-se por um momento, enquanto a mãe da Maria veste de novo a criança, e começa a tratar dos pedidos dos

<sup>1</sup> 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.

exames complementares de diagnóstico que pretende requisitar, tendo em conta que no topo da sua lista de diagnósticos diferenciais está um quadro de pneumonia. Faz o pedido de uma radiografia do tórax, face e perfil direito, e preenche o pedido de um hemograma, ionograma, bioquímica simples e determinação da proteína C reactiva (PCR), dado que são os testes que considera necessários e que habitualmente são pedidos na sua instituição nestes casos. Ao fazer estes pedidos tenta lembrar-se das razões porque considera necessários cada um dos testes que solicita, tornando-se claro no seu espírito que uma das principais questões a que tenta responder é a da etiologia vírica ou bacteriana do eventual quadro de pneumonia, dado que isto terá consequências extremamente importantes na decisão sobre o plano terapêutico. Em particular, relativamente à questão típica e fundamental da necessidade de prescrição ou não de antibióticos. Pondera sobre quais os dados mais úteis nesta distinção entre pneumonias bacterianas e víricas, para além da história clínica e do exame físico, que neste caso não eram muito esclarecedores. A radiografia torácica ajudará, assim como o hemograma (nível e tipo de leucocitose) e a PCR, no entanto, não pode deixar de suspirar ao pensar que nestas situações, não raras vezes, mesmo depois destes testes, continua com dúvidas. Neste exercício, acaba por lembrar-se que, na semana anterior, os técnicos do laboratório de bioquímica do hospital tinham estado, numa reunião de serviço, a falar sobre um teste que tinham implementado recentemente no hospital – a procalcitonina – e que era útil na distinção entre infecções víricas e bacterianas. Perante uma hipótese de reduzir a sua incerteza e ansiedade num problema com que se depara comumente na urgência, e porque nunca antes tinha ouvido falar da determinação da procalcitonina neste contexto, o interno decide, enquanto espera pelos resultados dos exames solicitados, procurar qual a evidência científica existente sobre a utilidade da determinação da procalcitonina na discriminação entre pneumonias de etiologia bacteriana e vírica na criança.

## FORMULAÇÃO DA QUESTÃO

Neste contexto, coloca-se ao interno uma questão relativa à utilidade, e em particular a validade, da utilização da determinação da procalcitonina no diagnóstico de pneumonia de etiologia bacteriana, e mais especificamente na distinção entre pneumonias de etiologia bacteriana e vírica. Assim, ao pensar na formulação da questão clínica, deverá ter em atenção os seus quatro elementos essenciais, seguindo o acrónimo que os identifica – PICO<sup>(5)</sup>: (1) a população são crianças com sinais e sintomas compatíveis com pneumonia; (2) a intervenção a avaliar será a determinação da procalcitonina e (3) a variável de resultado clínico será a determinação da existência e etiologia da pneumonia através de um teste (ou conjunto de testes) padrão de referência, neste caso, a identificação bacteriológica ou serológica do agente vírico ou bacteriano responsável pela infecção.

## PESQUISA DA EVIDÊNCIA

Tendo em conta a questão clínica de interesse, a procura de evidência científica deverá focar-se, essencialmente, em estudos sobre a validade do teste diagnóstico em avaliação – procalcitonina na distinção entre pneumonias de etiologia vírica ou bacteriana.

Neste caso, o interno começou por fazer uma pesquisa bibliográfica utilizando o serviço de acesso à base de dados *Medline* da *PubMed*<sup>(5)</sup>. Para isto, começou por identificar termos *MeSH*<sup>(5)</sup> associados ao termo “*procalcitonin*”, tendo verificado que este termo *MeSH* existe classificado como “*Substance name*” (“*procalcitonin*” [*Substance Name*]). Em seguida, decidiu utilizar, para a identificação do problema e patologia, os termos genéricos (*bacterial AND viral AND infection*). Adicionou termos para identificação do grupo etário de interesse (crianças, usando a definição dos 0 aos 18 anos de idade), com auxílio dos termos *MeSH* aplicados através da utilização da funcionalidade *Limits* do *Pubmed* (*infant [MeSH] OR child[MeSH] OR adolescent [MeSH]*). Por último, utilizando estas três partes da *query* ligadas por operadores booleanos *AND* utilizou a funcionalidade *clinical queries* do *Pubmed* para adicionar filtros de pesquisa metodo-

lógicos específicos para questões de diagnóstico, optando por um modo de pesquisa mais sensível (“*broad, sensitive search*”)<sup>(5)</sup>. Esta versão final da *query* de pesquisa (“*procalcitonin*” [*Substance Name*]) *AND* (*bacterial AND viral AND infection*) *AND* (*infant [MeSH] OR child[MeSH] OR adolescent [MeSH]*) *AND* (*Diagnosis/Broad[filter]*)) permitiu-lhe, à data de realização da pesquisa, a identificação de 40 artigos. Uma rápida análise dos títulos e resumos destes resultados permitiu-lhe identificar vários artigos altamente relevantes, sobre a questão em análise e no grupo etário pretendido. Na posição 7 da lista, encontrou 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*”<sup>(9)</sup>. Na verdade, este é só um dos vários estudos encontrados e que respondem à questão em análise, sendo a sua escolha ditada por ser um dos mais recentes e um dos que responde de forma mais específica à questão, após análise do seu resumo. Neste caso, poderia ser de enorme utilidade a existência de uma revisão sistemática e eventual estudo de meta-análise sobre esta matéria, no entanto, a discussão será desenvolvida relativamente à avaliação crítica deste artigo original em particular. O passo seguinte foi aceder à versão integral do artigo e iniciar a sua análise crítica.

## AVALIAÇÃO CRÍTICA DA EVIDÊNCIA – ESTUDOS SOBRE VALIDADE DE TESTES DIAGNÓSTICOS

O diagnóstico é um dos mais nobres actos da prática médica. 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, mas de uma forma mais geral pode ser contemplada a discriminação entre diferentes formas de uma doença ou, eventualmente, estádios diferentes de gravidade, progressão ou impacto da doença). 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 exame físico.

A utilização adequada e consciente dos testes diagnósticos exige o conhecimento e compreensão das propriedades fundamentais que os caracterizam. Destas, destacam-se as seguintes:

- (a) A reprodutibilidade (*reliability* ou *reproducibility*), definida como a concordância entre medições ou classificações repetidas de um mesmo fenómeno (intra-observador, intra-instrumento, inter-observadores ou inter-instrumentos). Esta é uma das propriedades mais importantes, e muitas vezes esquecida, de um teste diagnóstico, sendo obvio, por exemplo, que não faz sentido falar em validade de um teste se primeiro não for garantido que o mesmo é adequadamente reprodutível<sup>(10)</sup>. É importante sublinhar que para qualquer teste diagnóstico, e ao contrário do que é habitual fazer-se, é recomendada a prévia confirmação da sua reprodutibilidade, antes mesmo de olhar à sua validade.
- (b) A validade (*accuracy* ou *validity*) do teste diagnóstico, que será abaixo definida e abordada em detalhe ao longo deste artigo.
- (c) A caracterização da aceitabilidade, riscos e custos associados ao teste diagnóstico.
- (d) A caracterização dos efeitos da utilização do teste diagnóstico nas decisões e variáveis de resultado clínico relevantes e a sua custo-efectividade.

Sendo várias as suas características ou propriedades relevantes, é muito comum, no entanto, colocar uma ênfase especial na validade dos testes diagnósticos, 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 classificar adequadamente o indivíduo como não doente quando este é não doente e doente quando este é doente. Este estudo é feito 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. Tal como acontece noutras áreas da investigação clínica, quando pretendemos estudar a validade ou utilidade diagnóstica de um determinado teste diagnóstico existem várias fases ou questões de complexidade crescente que podem ser colocadas e cujas respostas exigem o adequado desenho de estudos com características, dificuldades e viéses específicos<sup>(11-13)</sup>. Neste contexto Sackett e Haynes definem quatro fases com questões distintas no âmbito do estudo da validade ou utilidade diagnóstica de um teste<sup>(11)</sup>:

#### Fase 1

Tenta responder à questão “será que os resultados do teste em indivíduos com a entidade nosológica em consideração (doentes) são diferentes dos resultados do teste em pessoas normais/saudáveis?”.

#### Fase 2

Tenta responder à questão “será que os indivíduos com determinados resultados do teste têm maior probabilidade de ter a doença do que indivíduos com outros resultados diferentes?”.

#### Fase 3

Tenta responder à questão “será que os resultados do teste diagnóstico

em avaliação conseguem discriminar adequadamente entre indivíduos com e sem a doença, no âmbito da população de indivíduos em quem faria sentido suspeitar da presença da mesma no contexto real da prática clínica onde o teste será potencialmente utilizado?”.

#### Fase 4

Tenta responder à questão “será que os indivíduos submetidos ao teste diagnóstico em avaliação acabam por atingir os resultados clínicos pretendidos (probabilidade e tempo de sobrevivência, capacidade funcional, qualidade de vida, etc.) de forma mais eficaz?”.

A discussão que se segue refere-se essencialmente aos estudos de fase 3, que são os mais habitualmente encontrados na literatura desta área.

O interno que atendeu a Maria no SU já encontrou o estudo sobre a validade da procalcitonina sérica na discriminação entre pneumonias de etiologia bacteriana e vírica que procurava. Assim, o desafio seguinte será o de definir e aplicar um conjunto de critérios que lhe permitam avaliar este artigo relativamente a três aspectos fundamentais (Tabela 1): (A) a qualidade metodológica do estudo; (B) a importância dos seus resultados e (C) a aplicabilidade prática dos mesmos. No presente artigo será desenvolvida a discussão dos aspectos relacionados com a avaliação da qualidade metodológica do estudo e no número seguinte desta série discutir-se-ão as questões relativas à avaliação quantitativa e qualitativa da importância dos resultados e a sua aplicabilidade na prática clínica.

#### (A) AVALIAÇÃO DA QUALIDADE METODOLÓGICA DO ESTUDO

O estudo da validade de um teste diagnóstico faz-se, habitualmente, através da comparação concorrente e independente do seu resultado com o resultado de um teste padrão (*gold standard*), assumido para determinada situação clínica e num determinado momento no tempo. Desta forma, o tipo de estudos geralmente encontrados, e mais apropriados neste âmbito, são estudos observacionais, tipicamente transversais (pros-

**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 otimizar 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?
- 

pectivos ou retrospectivos), com aplicação concorrente e independente do teste em avaliação e do teste padrão numa amostra apropriada de indivíduos da população em estudo. Neste contexto, e tendo em atenção as várias fontes de erros sistemáticos e aleatórios que podem surgir neste tipo de estudos<sup>(12,13)</sup>, definem-se habitualmente os seguintes critérios (Tabela 1) para avaliação da qualidade metodológica de estudos sobre a validade de testes diagnósticos<sup>(4,12-18)</sup>:

**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?**

Diferenças sócio-demográficas e clínicas entre populações podem levar a que as estimativas de validade de um determinado teste diagnóstico variem bastante. Este fenómeno deve-se mais à natural heterogeneidade destas medidas e à impossibilidade de generalização dos resultados do estudo do que à existência de reais erros sistemáticos ou viéses.

As estimativas da validade dos testes diagnósticos poderão ter utilidade clínica muito limitada se o espectro de doença nos participantes seleccionados para o estudo não é representativo daquele encontrado na população que irá, potencialmente, ser submetida ao teste no contexto clínico real de utilização do mesmo. O termo espectro de doença refere-se não só à gravidade e evolução da entidade nosológica em consideração, mas também às características sócio-demográficas dos indivíduos e à presença e frequência de outros diagnósticos diferenciais e/ou comorbilidades. Os estudos mais úteis são aqueles que incluem participantes que reflectem adequadamente a diversidade encontrada no contexto prático real. Estudos que comparam indivíduos nos extremos do espectro da doença (ex: comparações de indivíduos definitivamente doentes com indivíduos saudáveis) tendem a sobrestimar a validade e utilidade dos testes diagnósticos<sup>(11,19,20)</sup>. É fundamental que o estudo inclua um espectro adequado de participantes e que os critérios e métodos de selecção

dos mesmos sejam descritos de forma adequada e completa.

No estudo de Don et al.<sup>(9)</sup>, cuja análise crítica agora iniciamos, verifica-se que os autores descrevem na subsecção “*Study subjects*” da secção de métodos que recrutaram consecutivamente 125 crianças previamente saudáveis e com suspeita clínica de pneumonia, no contexto de um serviço de urgência hospitalar numa cidade italiana. Os autores excluíram 25 das crianças por não terem amostras sanguíneas (n=17) ou exames radiográficos não compatíveis com pneumonia (n=8). As 100 crianças finalmente incluídas no estudo tinham uma mediana de idades de 3,7 anos, 19% tinham < 24 meses, 38% tinham ≥ 5 anos e 49% eram do sexo masculino. Destes 100 doentes com pneumonia confirmada, 74 foram tratados em ambulatório e 26 necessitaram de internamento hospitalar; 18 tinham pneumonia de etiologia pneumocócica, 25 bacteriana atípica (*Mycoplasma* + *Clamídia*), 23 vírica e 34 eram de etiologia indeterminada. A descrição feita pelos autores sugere que a amostra estudada seja adequadamente representativa da população de doentes a quem se aplicaria o teste da procalcitonina na prática clínica em contexto real para a discriminação entre pneumonias de etiologia bacteriana e vírica.

## **2) Os critérios de selecção dos participantes, o contexto, o local e o período em que o estudo foi realizado são claramente descritos?**

Os testes diagnósticos podem ter propriedades distintas em função do espectro de doença, características sócio-demográficas, presença e frequência de diagnósticos diferenciais e/ou comorbilidades, contexto de aplicação do teste (cuidados primários de saúde versus cuidados secundários ou terciários) ou objectivo da sua aplicação (rastreamento versus confirmação diagnóstica). Por isto, e em íntima ligação com o descrito no ponto anterior, os estudos sobre validade de testes diagnósticos deverão conter uma descrição adequada e completa dos critérios de selecção (inclusão e exclusão) dos participantes, do contexto e local nos quais o estudo foi realizado e do período temporal de recrutamento considera-

do. Esta informação é fundamental para poder fazer-se um adequado julgamento sobre o espectro de participantes estudados e sobre a generalização possível dos resultados do estudo.<sup>(14,19,20)</sup>

No estudo de Don et al.<sup>(9)</sup>, na secção de métodos, tal como mencionado no ponto anterior, são referidos adequadamente os critérios de selecção dos participantes<sup>(21)</sup>, o contexto, o local e o período em que o estudo foi realizado. É, assim, possível confirmar a adequação e representatividade da amostra de crianças analisada. Os critérios de inclusão e exclusão utilizados sugerem que os resultados do estudo, se válidos, possam ser generalizáveis para crianças sem patologias crónicas com suspeita de pneumonia vistas em serviços de urgência de qualquer país europeu, nomeadamente, generalizáveis para o nosso contexto português.

## **3) Os métodos de recrutamento/amostragem dos participantes são claramente descritos? Foi usado um método de amostragem consecutiva seguindo os critérios de selecção estabelecidos?**

Um outro elemento fundamental neste tipo de estudos é o método de recrutamento dos participantes. Existem vários possíveis métodos de recrutamento neste tipo de estudos<sup>(19)</sup>. Um dos métodos mais típico e adequado é o recrutamento consecutivo de indivíduos apresentando sinais e/ou sintomas sugestivos ou compatíveis com a entidade nosológica em consideração ou referenciados por outros profissionais de saúde com esta suspeita. Outros métodos de recrutamento possíveis, eventualmente menos apropriados, incluem<sup>(11,19)</sup>: (a) recrutamento de indivíduos após aplicação do teste de referência e com aplicação subsequente do teste padrão; (b) recrutamento após aplicação do teste padrão de referência e confirmação da presença ou ausência da doença, com aplicação subsequente do teste em avaliação; (c) recrutamento de indivíduos com confirmação prévia da presença da doença (casos) e recrutamento independente de indivíduos sem doença (controles); (d) estudos retrospectivos baseados em registos hospitalares e com identificação

de indivíduos a quem foi aplicado o teste em avaliação, o teste padrão ou ambos. Quanto ao processo de amostragem, os métodos mais apropriados são a amostragem consecutiva, sistemática ou aleatória. Obviamente, estas diferentes opções de recrutamento/amostragem de participantes terão implicações directas nas características dos mesmos, no espectro de doença e na frequência de diagnósticos alternativos e comorbilidades. Assim, uma descrição clara do desenho do estudo e métodos de recrutamento e amostragem é fundamental na avaliação crítica deste tipo de estudos.

No estudo de Don et al.<sup>(9)</sup>, na secção de métodos, tal como mencionado no ponto anterior, são descritos adequadamente e de uma forma completa o desenho do estudo e os métodos de amostragem e recrutamento dos participantes<sup>(21)</sup>. Foi feito o recrutamento de uma amostra consecutiva de 100 crianças com suspeita de pneumonia seguindo os critérios de selecção previamente definidos.

## **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)?**

O planeamento do estudo e da recolha de dados previamente à realização dos testes envolvidos (estudos prospectivos) permite um controlo mais adequado dos dados e a obtenção de dados mais completos e apropriados, garantindo uma qualidade muito superior dos mesmos e do estudo em geral (menor quantidade de variáveis em falta e de dados omissos ou ininterpretáveis).<sup>(14)</sup>

Os estudos planeados após a realização dos testes em avaliação e dos testes padrão de referência (estudos retrospectivos), geralmente dependem da consulta e recolha de informação a partir de registos clínicos e/ou administrativos, e por esta razão estão geralmente associados a dados de menor qualidade, menos completos, eventual enviesamento na selecção e recrutamento dos participantes e eventual violação do princípio da independência entre os resultados dos testes comparados.<sup>(13,14,20)</sup>

No estudo de Don et al.<sup>(9)</sup>, a descrição dos métodos utilizados torna claro que o estudo foi planeado previamente, sendo um estudo transversal e prospectivo com medição concorrente do teste em avaliação (procalcitonina) e do teste padrão de referência (diagnóstico clínico, radiológico e confirmação serológica da etiologia da pneumonia), utilizando um protocolo previamente definido.

**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?**

Um teste padrão de referência (*gold standard*) é genericamente definido como o teste ou conjunto de testes que, num determinado momento no tempo, é aceite consensualmente como sendo o que melhor consegue reflectir o verdadeiro estado de doença ou não doença do indivíduo, no contexto em consideração. É a partir da comparação entre os resultados deste e do teste em avaliação que podemos estimar as suas medidas de validade. As metodologias habituais para a estimação da validade de um teste diagnóstico assumem implicitamente que o teste padrão de referência tem uma validade perfeita (100% de sensibilidade e de especificidade) e que, portanto, quaisquer discrepâncias entre este e o teste em avaliação são fruto das deficiências deste último. No entanto, dever-se-á ter em atenção que (1) em muitas circunstâncias não existem, de facto, testes padrão de referência; (2) muitas vezes existe discordância entre os especialistas quanto ao padrão mais adequado; (3) o padrão disponível pode não ter validade perfeita (e pode até, eventualmente, não ser melhor que o teste em avaliação) e (4) o teste padrão pode implicar riscos ou ser demasiado invasivo, sendo necessário substituí-lo por alternativas menos invasivas. Assim sendo, a escolha do padrão de referência é um aspecto de extrema importância e, por vezes, difícil. O julgamento sobre a adequação do teste padrão de referência é um aspecto fundamental na avaliação da qualidade deste tipo de estudos e exige conhecimentos e experiência específicos na área clínica em consideração.<sup>(13,14,18)</sup>

No estudo de Don et al.<sup>(9)</sup> interessava dispor de um teste padrão de referência que, em doentes com sinais e sintomas compatíveis com pneumonia, distinguísse entre etiologias bacterianas e víricas. Os testes padrão utilizados pelos autores e descritos na subsecção de "*Aetiological classification*" dos métodos foram os testes serológicos para detecção de 5 tipos diferentes de bactérias e 9 tipos de vírus mais frequentes<sup>(21)</sup>. Genericamente estes testes padrão de referência poderiam ser potencialmente adequados ao diagnóstico da etiologia da pneumonia, no entanto, deverá ser sublinhado que em 34 dos 100 participantes não foi possível, com a bateria utilizada, identificar especificamente o agente causador da pneumonia. Este facto torna difícil a interpretação dos resultados e pode condicionar potenciais vieses associados a sobre ou subestimação das medidas de validade do teste em avaliação. Apesar de serem explícita e adequadamente descritos, estes casos de etiologia desconhecida têm importantes implicações e pela sua magnitude (34%) limitam a utilidade dos resultados do estudo.

**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?**

Idealmente a aplicação do teste em avaliação e do teste padrão deverão ser feitas concorrentemente. Quando não é possível obter simultaneamente os resultados de ambos os testes existe sempre um risco de má classificação, crescente com a magnitude do período temporal, devido à eventual alteração do estado do indivíduo ou a sua doença. Os períodos temporais que poderão de facto levar à existência deste tipo de vieses são variáveis em função do tipo de doença ou situação clínica.<sup>(13,14,18)</sup>

No estudo de Don et al.<sup>(9)</sup>, o protocolo de recolha de dados implicava a colheita de sangue periférico segundo rotina habitual na altura da admissão no serviço de urgência e este servia para a determinação quer do teste em avalia-

ção (procalcitonina) quer da bateria de testes serológicos usados como padrão de referência. A realização dos testes propriamente ditos foi feita nos dias que se seguiram à colheita, não havendo um espaço temporal significativo entre a aplicação do teste e do padrão.

**7) Todos os participantes, ou uma amostra aleatória dos participantes, foram submetidos ao teste padrão de referência?**

Idealmente todos os participantes deveriam ser submetidos ao teste em avaliação e ao teste padrão de referência. Na impossibilidade de serem todos submetidos a ambos os testes, uma alternativa aceitável é a de submeter ao teste padrão apenas uma amostra aleatória dos participantes, garantindo a representatividade e não enviesamento das estimativas de validade. Infelizmente, em muitas circunstâncias verifica-se que os testes padrão de referência só são aplicados numa amostra dos participantes que é seleccionada por conveniência ou na dependência do resultado do teste em avaliação. Estas opções levam a estimativas enviesadas da validade do teste e deverão ser evitadas, estando associadas ao chamado viés de verificação parcial (*partial verification bias*, *work-up bias* ou *sequential ordering bias*).<sup>(13,18)</sup>

No estudo de Don et al.<sup>(9)</sup>, todos os doentes foram submetidos ao teste padrão de referência, no entanto, é de salientar que em 34 dos 100 doentes não foi possível identificar o agente causador da patologia, levantando a hipótese destas etiologias desconhecidas poderem condicionar enviesamento na estimativa das medidas de validade do teste.

**8) Todos os participantes receberam o mesmo teste padrão de referência independentemente do resultado do teste diagnóstico em avaliação?**

O chamado viés de verificação diferencial (*differential verification bias*) acontece quando para alguns dos participantes do estudo os resultados do teste em avaliação são verificados/comparados com os resultados de testes padrão de referência diferentes daqueles aplicados

nos restantes participantes. Isto é particularmente problemático se os testes padrão utilizados são diferentes quanto à definição ou capacidade de discriminação da entidade nosológica em consideração ou quando a aplicação dos diferentes padrões está associado ao resultado do próprio teste a avaliar (ex: utilização de um padrão mais invasivo e mais válido em doentes que tenham resultados positivos no teste em avaliação). Estas situações estão, naturalmente, associadas a enviesamento na determinação da validade do teste.<sup>(13,14,18)</sup>

No estudo de Don et al.<sup>(9)</sup>, todos os participantes receberam o mesmo conjunto de testes padrão seguindo um protocolo comum e independentemente do resultado do teste 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)?

Quando o próprio teste em avaliação é utilizado para estabelecer o diagnóstico final, ou seja, quando este faz parte integrante do padrão de referência utilizado, as medidas de validade obtidas serão, muito provavelmente, enviesadas (*incorporation bias*).<sup>(13,18)</sup> Este aspecto só é aplicável quando o teste padrão de referência é constituído por um conjunto de testes, em vez de um teste único, e o resultado do teste em avaliação, para além de ser conhecido, é incorporado de facto nesse conjunto.

No estudo de Don et al.<sup>(9)</sup>, que temos vindo a avaliar criticamente, quer os critérios de selecção quer o teste padrão de referência são independentes do teste em avaliação (procalcitonina).

## 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?

A forma de aplicação e execução do teste padrão de referência e do teste em avaliação deverão ser descritas de forma detalhada e completa ou dadas referências adequadas para o efeito. Este aspecto é importante para controlar a varia-

bilidade que poderá surgir na sequência de protocolos distintos de aplicação dos testes e para garantir a possibilidade de replicação dos mesmos noutros contextos.<sup>(14,18)</sup>

No estudo de Don et al.<sup>(9)</sup>, nas subsecções de “*Aetiological classification*” e “*Laboratory studies*” da secção de métodos são descritos em detalhe e devidamente referenciados os protocolos de execução técnica dos testes padrão de referência e do teste em avaliação.

## 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)?

Os resultados do teste padrão e do teste em avaliação deverão ser obtidos e interpretados sem o conhecimento do resultado um do outro. A ocultação dos resultados dos testes é importante uma vez que o conhecimento prévio dos resultados poderá levar a estimativas enviesadas da validade do teste em avaliação (*review bias*). A importância deste aspecto poderá ser tanto maior quanto mais afectada de subjectividade estiver a obtenção e interpretação dos resultados dos testes.<sup>(13,14,18)</sup>

No estudo de Don et al.<sup>(9)</sup>, na secção de métodos é descrito o protocolo de execução dos testes padrão e do teste em avaliação, sendo possível verificar que as execuções dos mesmos foram feitas em laboratórios independentes e sem conhecimento prévio dos resultados uns dos outros.

## 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?

A obtenção e interpretação dos resultados de testes diagnósticos podem ser afectadas pela existência, nesse momento, de outra informação clínica relevante. Assim, e para que as estimativas

das medidas de validade de um teste diagnóstico não sejam enviesadas, o ideal será que a informação disponível na altura da obtenção e interpretação do resultado do teste seja idêntica àquela que estará disponível no contexto clínico real de utilização do mesmo.<sup>(14,18)</sup>

No estudo de Don et al.<sup>(9)</sup>, o protocolo de selecção de participantes e de recolha de dados reproduz de forma bastante próxima aquilo que aconteceria a uma criança com suspeita de pneumonia que recorre a um serviço de urgência num contexto clínico real. Desta forma, é seguro assumir que a informação clínica disponível era semelhante àquela que está habitualmente disponível no contexto clínico real de utilização do teste.

## 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?

Não raramente os testes diagnósticos têm resultados classificados como ambíguos, intermédios ou não interpretáveis. Muitas vezes os investigadores optam por ignorar na análise e não reportar estes resultados, sendo que isto pode levar ao enviesamento das estimativas de validade dos testes diagnósticos, especialmente se, por qualquer razão, estes resultados estão associados ao verdadeiro estado do indivíduo (doente ou não doente). Frequentemente, também, existem situações em que por uma ou outra razão os resultados dos testes não estão disponíveis (dados omissos). Da mesma forma, estas situações podem levar ao enviesamento das estimativas de validade, especialmente quando associadas ao verdadeiro estado do indivíduo. Idealmente, os investigadores deverão sempre reportar este tipo de situações e as eventuais razões pelas quais estas surgiram, de forma a que o leitor possa julgar o impacto real da sua existência nas estimativas de validade do teste.<sup>(13,14,18)</sup>

No estudo de Don et al.<sup>(9)</sup> é particularmente relevante a existência de 34% de testes padrão de referência classificados como tendo “etiologia desconhecida”.

Os resultados destes testes padrão não interpretáveis são adequada e explicitamente descritos neste artigo, no entanto, podem condicionar potenciais vieses na estimação das medidas de validade. Relativamente ao teste em avaliação (procalcitonina), todos os 100 participantes finais do estudo tiveram a sua determinação realizada e devidamente reportada. Chama-se, no entanto, a atenção para a existência de 17 crianças das 125 inicialmente recrutadas para as quais não foi possível obter amostras de sangue que permitissem a realização dos testes. Esta situação é devidamente reportada no artigo e os indivíduos foram excluídos da análise.

**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?**

O abandono do estudo ou a perda de seguimento, no caso de estudos prospectivos, são fenómenos importantes e por vezes de magnitude preocupante. Estes fenómenos podem enviesar de forma importante as estimativas de validade, especialmente quando a sua ocorrência não acontece de forma aleatória, mas antes associada ao verdadeiro estado de doente ou não doente dos participantes. Este enviesamento existe porque os indivíduos que abandonam ou são perdidos são tipicamente diferentes daqueles que permanecem no estudo. Razões frequentes para este tipo de ocorrências são os eventos adversos, a invasividade dos testes ou o estado de saúde (eventualmente diminuído) dos indivíduos, estando, por exemplo, estas razões associadas geralmente à presença de doença, e por isso ligadas a um eventual enviesamento das medidas de validade.<sup>(13,14,18)</sup>

No estudo de Don et al.<sup>(9)</sup>, não existe registo de quaisquer abandonos ou perdas de seguimento nos participantes do estudo, não constituindo este último ponto um problema para a qualidade metodológica do artigo em avaliação.

**(B) E (C)  
AVALIAÇÃO DA IMPORTÂNCIA E DA APLICABILIDADE PRÁTICA DA EVIDÊNCIA**

A avaliação da importância científica e prática dos resultados passa pela adequada interpretação e contextualização das medidas de validade do teste diagnóstico (sensibilidade, especificidade, valores preditivos, *likelihood ratios*, etc.) e subsequente verificação, em última análise, da real capacidade de discriminação do teste diagnóstico entre os indivíduos com e sem a entidade nosológica em consideração. A definição, cálculo e demonstração da interpretação e relevância destes conceitos serão discutidos no próximo número desta série.

A última questão no processo de avaliação crítica da evidência será a questão da aplicabilidade prática da mesma. 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. Na segunda parte deste artigo, serão também exploradas as questões e critérios aplicáveis à avaliação da aplicabilidade prática dos resultados deste tipo de estudos.

**CONCLUSÃO**

O diagnóstico é um dos actos mais nobres da prática clínica. A utilização adequada dos testes diagnósticos depende do conhecimento das suas propriedades e características operacionais, em particular da sua validade, e do desenvolvimento de competências que permitam maximizar a utilização da informação que a partir deles é possível obter. Ao longo deste artigo 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 um conjunto de critérios básicos para a avaliação da qualidade metodológica destes estudos foi discutida e ilustrada através de um exemplo prático. Na segunda parte deste artigo, serão discutidos a aplicação prática das medidas de validade dos testes

diagnósticos e os princípios quantitativos que permitem integrar a informação dada pelo resultado de um teste no processo de decisão diagnóstica.

O estudo de Don et al.<sup>(9)</sup> sobre a validade da procalcitonina na discriminação entre pneumonias de etiologia bacteriana e vírica na criança foi apresentado e discutido ao longo deste artigo. Verificou-se que este artigo, apesar de globalmente adequado quanto à sua qualidade metodológica, tem dois problemas importantes: (1) a utilização de um teste padrão de referência que, sendo genericamente adequado, não consegue identificar a etiologia específica em 34% dos doentes, com o potencial de enviesamento associado e a consequente redução da amostra a analisar e (2) o estudo tem um tamanho de amostra limitado (agravado pelos resultados inconclusivos do padrão de referência), com consequentes estimativas de validade do teste diagnóstico de precisão muito limitada (erros aleatórios associados às pequenas amostras) e, para além disto, os autores não reportam explicitamente a precisão das suas estimativas, não dando oportunidade aos leitores de avaliar criticamente este aspecto.

No próximo número desta série será discutido o desfecho do cenário clínico apresentado e, à luz da avaliação quantitativa e qualitativa da importância dos resultados do estudo e da aplicabilidade prática dos mesmos, dando continuidade à avaliação crítica do artigo de Don et al.<sup>(9)</sup>, será discutida a relevância e adequação da utilização da procalcitonina na discriminação entre pneumonias de etiologia vírica e bacteriana na criança.

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

**ABSTRACT**

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 a practical example is developed allowing the application of concepts, methods and

competencies covered in previous articles and specifically devoted to the critical appraisal and practical implementation of diagnostic tests accuracy studies. A clinical scenario is presented and explored regarding the search and critical appraisal of scientific evidence about the accuracy of serum procalcitonin in the distinction of bacterial from viral pneumonia in children. Methods for searching the current best evidence are discussed and a systematic approach for the critical appraisal of diagnostic tests accuracy studies is 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. In the present article we will discuss the search and critical appraisal process covering only the assessment of the methodological quality of the study, in the next article of this section matters regarding the quantitative and qualitative assessment of the impact of results and their practical applicability will be covered.

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

Nascer e Crescer 2010; 19(2): 116-24

## BIBLIOGRAFIA

1. Azevedo LF, Costa Pereira A: Pediatria Baseada na Evidência. *Nascer e Crescer Revista do Hospital de Crianças Maria Pia* 2007, vol 16:29-31.
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, 268:2420-2425.
3. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS: Evidence based medicine: what it is and what it isn't. *Bmj* 1996, 312:71-72.
4. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB: *Evidence-based medicine : how to practice and teach EBM* 2nd edn. 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, 16:135-140.
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, 17:30-36.
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, 17:30-38.
8. Azevedo LF, da Costa Pereira A: Pediatria Baseada na Evidência - Avaliação Crítica e Implementação Prática de Ensaio Clínicos Aleatorizados. *Nascer e Crescer - Revista do Hospital de Crianças Maria Pia* 2009, 18:106-119.
9. Don M, Valent F, Korppi M, Falletti 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, 39:129-137.
10. Lucas NP, Macaskill P, Irwig L, Bogduk N: The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol* 2010.
11. Sackett DL, Haynes RB: The architecture of diagnostic research. *Bmj* 2002, 324:539-541.
12. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH, Bossuyt PM: Empirical evidence of design-related bias in studies of diagnostic tests. *Jama* 1999, 282:1061-1066.
13. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J: Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004, 140:189-202.
14. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Rennie D, de Vet HC, Lijmer JG: The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003, 138:W1-12.
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, 271:389-391.
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, 271:703-707.
17. Whiting P, Rutjes AW, Dinnes J, Reitsma J, Bossuyt PM, Kleijnen J: Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess* 2004, 8:iii, 1-234.
18. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J: The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003, 3:25.
19. Knottnerus JA, Muris JW: Assessment of the accuracy of diagnostic tests: the cross-sectional study. *J Clin Epidemiol* 2003, 56:1118-1128.
20. van der Schouw YT, Van Dijk R, Verbeek AL: Problems in selecting the adequate patient population from existing data files for assessment studies of new diagnostic tests. *J Clin Epidemiol* 1995, 48:417-422.
21. Don M, Fasoli L, Paldanius M, Vainionpaa R, Kleemola M, Raty R, Leinonen M, Korppi M, Tenore A, Canciani M: Aetiology of community-acquired pneumonia: serological results of a paediatric survey. *Scand J Infect Dis* 2005, 37:806-812.

## CORRESPONDÊNCIA

lazevedo@med.up.pt.