

# Avaliação Crítica e Implementação Prática de Ensaio Clínicos Aleatorizados

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

## 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 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 ensaios clínicos aleatorizados (ECA's). É apresentado um cenário clínico em que surge a necessidade de encontrar e avaliar a evidência científica existente sobre a eficácia da utilização de antibióticos no tratamento da otite média aguda em crianças com menos de 2 anos. São discutidos os métodos de pesquisa da evidência mais actual e mais apropriada à questão; e é sugerida uma metodologia sistemática para a avaliação crítica de ECA's incluindo três fases distintas: (1) avaliação da validade ou qualidade metodológica do ECA; (2) avaliação da importância científica e prática dos seus resultados e (3) avaliação da aplicabilidade prática dos mesmos.

Palavras-chave: Ensaio clínico aleatorizado, medicina baseada na evidência, otite média aguda, antibióticos, observação expectante.

Nascer e Crescer 2009; 18(2): 106-119

<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.

## 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 (validade ou qualidade metodológica, importância científica e prática dos resultados e aplicabilidade prática dos mesmos). No quarto artigo<sup>(7)</sup> foi iniciada a apresentação de um conjunto de exemplos e 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.

Neste quinto artigo da série será apresentado um cenário prático no qual se exemplifica a aplicação destes conceitos, métodos e competências no âmbito da avaliação crítica e implementação prática de ensaios clínicos aleatorizados.

## CENÁRIO CLÍNICO

Manuel, de 18 meses de idade, é trazido pela sua mãe ao consultório do seu pediatra assistente com febre, há cerca de 12 horas, e depois de ter estado acordado, choroso e irritado durante a noite. O Manuel ainda consegue manter uma hidratação adequada e não está prostrado, no entanto, é visível que tanto ele como a sua mãe estão exaustos devido à noite mal dormida. No exame físico a única alteração relevante é a existência

de alterações na otoscopia, com rubor e ligeiro abaulamento da membrana timpânica à direita. O Manuel é uma criança sem antecedentes patológicos ou familiares relevantes e sem história de otites prévias. Não apresenta história de alergias a antibióticos ou outros medicamentos. O diagnóstico do seu pediatra assistente é de otite média aguda (OMA).

A mãe do Manuel, exausta e ligeiramente preocupada, solicita ao pediatra a prescrição de um antibiótico, referindo que esse é o único tratamento que funciona no seu sobrinho de 5 anos que está constantemente a ter otites.

O pediatra do Manuel assistiu recentemente a uma sessão de formação sobre utilização racional de antibióticos e prevenção de resistências, e está consciente da necessidade de evitar, tanto quanto possível, a utilização excessiva de antibióticos. Adicionalmente, lembra-se de um colega pediatra mais experiente lhe ter referido, durante a semana anterior, que não utiliza por rotina antibióticos no tratamento de otites médias agudas (OMA), facto que o deixou algo surpreendido. Neste cenário, o pediatra assistente do Manuel questiona-se sobre a necessidade de instituir terapêutica antibiótica neste caso, e decide procurar evidência científica que lhe permita apoiar a sua decisão.

## FORMULAÇÃO DA QUESTÃO

Neste contexto, coloca-se ao pediatra uma questão relativa à eficácia da terapêutica antibiótica no tratamento da otite média aguda (OMA) numa criança com menos de dois anos. Assim, ao pensar na formulação de questão clínica, dever-se-ão ter em atenção os seus quatro elementos essenciais<sup>(8)</sup>: (1) a população são crianças com menos de dois anos, sem antecedentes relevantes, com

um quadro de OMA sem outras complicações; (2) a intervenção a avaliar será a terapêutica antibiótica; (3) a intervenção alternativa será o tratamento placebo e/ou a observação expectante e (4) as variáveis de resultado clínico relevantes neste caso dizem respeito à resolução clínica e à prevenção de complicações a curto, médio e longo prazo.

## PESQUISA DA EVIDÊNCIA

Tendo em conta que a questão clínica de interesse é a da eficácia de uma intervenção terapêutica, a procura de evidência científica deverá focar, essencialmente, nos ensaios clínicos aleatorizados (ECA), pois estes são os estudos primários com maior nível de evidência científica neste tipo de questões.

Neste caso, o pediatra começou então por fazer uma pesquisa bibliográfica utilizando o serviço de acesso à base de dados *Medline* da *PubMed*<sup>(5)</sup>, com o objectivo de encontrar ECA's sobre a eficácia da terapêutica antibiótica no tratamento da OMA, em crianças com menos de dois anos. Para isto, começou por identificar os termos *MeSH*<sup>(6)</sup> associados a *acute otitis media*, tendo verificado que só existe o termo *MeSH* "otitis media". Em seguida, decidiu utilizar, para a identificação da patologia, a seguinte *query* (*acute AND "Otitis Media"[Mesh]*). Adicionou termos para identificação do grupo etário de interesse (crianças com menos de 2 anos) e exclusão de outros grupos etários, com auxílio dos termos *MeSH* aplicados através da utilização da funcionalidade *Limits* do *Pubmed*. Assim, a segunda parte da *query* utilizada foi ("*Infant*"[*Mesh:NoExp*]NOT ("*child*"[*MeSH Terms:noexp*] OR "*adolescent*"[*MeSH Terms*] OR "*adult*"[*MeSH Terms:noexp*])). Adicionou também um termo para a identificação da intervenção em causa – *antib\** (a utilização do asterisco – *wildcard character* – permite incluir na pesquisa todos os possíveis termos que tenham a raiz comum identificada antes do mesmo, neste caso, serão incluídos por exemplo, *antibiotic*, *antibiotics*, *antibacterial*, *antibiotherapy*, etc). Utilizando estas três partes da *query* ligadas por operadores booleanos *AND*, aplicou a funcionalidade *clinical queries* da *Pubmed* para adicio-

nar filtros de pesquisa metodológicos específicos para questões de tratamento.

Esta versão final da *query* de pesquisa permitiu-lhe, à data de realização da pesquisa, a identificação de 40 artigos. Uma rápida análise dos títulos destes resultados permitiu-lhe identificar pelo menos um artigo altamente relevante, sobre a questão em análise e no grupo etário pretendido. Na posição 17 da lista, encontrou o artigo de *Damoiseau et al.*, publicado em 2000, no *British Medical Journal* e intitulado "*Primary care based randomised, double blind trial of amoxicillin versus placebo for acute otitis media in children aged under 2 years*"<sup>(8)</sup>. O passo seguinte foi aceder à versão integral do artigo e iniciar a sua análise crítica.

## AVALIAÇÃO CRÍTICA DA EVIDÊNCIA – ENSAIOS CLÍNICOS ALEATORIZADOS

Os ensaios clínicos aleatorizados são o exemplo paradigmático de estudos experimentais em investigação clínica. Estes distinguem-se dos estudos observacionais pela existência de um controlo directo da intervenção a avaliar e pela existência de uma atribuição aleatória da mesma pelos participantes no estudo (aleatorização ou randomização).

Pela sua natureza controlada e pelo uso da aleatorização, os ECA's são considerados como o método padrão de avaliação da eficácia de intervenções terapêuticas, sendo hoje uma parte fundamental do processo regulamentar de obtenção da autorização de introdução no mercado de qualquer medicamento ou dispositivo médico, e essenciais para justificar a utilização de qualquer intervenção terapêutica ou preventiva de uma forma geral. Neste âmbito, a discussão será feita em particular sobre os designados ensaios clínicos de fase III.

Tendo encontrado o ensaio clínico aleatorizado que pretendia, o desafio seguinte que se coloca ao pediatra assistente do *Manuel* será, então, definir e aplicar um conjunto de critérios que lhe permitam avaliar este artigo relativamente a três aspectos fundamentais (Tabela 1): (A) a sua validade ou qualidade metodológica; (B) a importância dos seus resultados e (C) a aplicabilidade prática dos mesmos.

## (A) Avaliação da validade ou qualidade metodológica

Apesar dos ECA's serem o padrão metodológico na resposta a questões sobre a eficácia de intervenções terapêuticas, devido ao seu carácter controlado e à aplicação da aleatorização, a verdade é que a qualidade destes está, tal como qualquer outro tipo de estudo, dependente do controlo de um vasto conjunto de outras importantes fontes de erros sistemáticos e aleatórios que devem ser conhecidas, detectadas e prevenidas. Neste contexto, definem-se habitualmente os seguintes critérios (Tabela 1) para avaliação da validade ou qualidade metodológica de ensaios clínicos aleatorizados<sup>(9-15)</sup>:

### 1) Houve uma adequada atribuição aleatória das intervenções (aleatorização ou randomização)?

O processo de atribuição aleatória das intervenções – aleatorização – é um pilar fundamental dos estudos experimentais em humanos, garantindo o controlo de potenciais erros sistemáticos decorrentes do confundimento e de vieses de selecção<sup>(9-12)</sup>. Esta é a vantagem fundamental dos ECA's relativamente aos estudos observacionais. Nestes segundos, a potencial existência de confundimento e vieses de selecção, por não haver aleatorização das intervenções ou factores em estudo, são barreiras formais à interpretação causal directa dos efeitos observados<sup>(9-12)</sup>.

Num ECA, o processo de atribuição aleatória da intervenção, se adequadamente executado, permite criar grupos (experimental e de controlo) comparáveis e homogéneos quanto às características ou variáveis que podem potencialmente influenciar o resultado clínico subsequente, sendo as diferenças entre estes grupos devidas unicamente ao processo aleatório que os determina. Por este motivo, é lícito interpretar as diferenças observadas prospectivamente entre estes grupos, quanto às variáveis de resultado clínico relevantes, como sendo resultantes do efeito causal da intervenção, desde que essas diferenças estejam para além das que seriam de esperar pelo mero acaso. A decisão sobre se as diferenças estão ou não além das esperadas decorre da

aplicação de testes estatísticos de hipóteses adequados, que permitem decidir, precisamente, se as diferenças são ou não estatisticamente significativas.

Do ponto de vista prático, a verificação da existência de um método adequado de atribuição aleatória das intervenções passa pela confirmação de que o processo utilizado foi realmente aleatório. A expressão “atribuição aleatória da intervenção” tem um significado técnico preciso, e implica que cada participante tenha uma probabilidade conhecida de pertencer a um ou outro grupo, definida por um processo aleatório; sendo, no entanto, impossível prever, à partida, a intervenção a que cada participante será submetido. Na generalidade dos casos, hoje em dia, este processo é feito através

de aplicações informáticas específicas, que utilizam algoritmos que garantem um processo aleatório de alocação das intervenções. Frequentemente, o processo é coordenado por centros de aleatorização, que remotamente e de forma independente controlam e garantem a qualidade da aleatorização. Continuam a encontrar-se, por vezes, métodos mais simples, mas quando bem usados igualmente eficazes (se bem que são de fácil manipulação e consequente risco de uso inadequado), como por exemplo a utilização de tabelas de números aleatórios ou outros métodos tradicionais de geração de resultados aleatórios (ex: a moeda ao ar)<sup>(9-12)</sup>.

A avaliação do processo de aleatorização deverá incluir também a verificação da existência de aleatorização estra-

tificada e/ou aleatorização permutada em blocos (incluindo o tipo e tamanho dos blocos utilizados)<sup>(9)</sup>.

A aleatorização estratificada resulta da aplicação de um processo independente de aleatorização das intervenções em determinados subgrupos de participantes, criados em função de variáveis especialmente relevantes no contexto do estudo<sup>(9,16)</sup>. Por exemplo, fazer uma aleatorização estratificada em função do sexo implica desenvolver o processo de atribuição aleatória da intervenção, independentemente, no subgrupo de participantes do sexo feminino e no subgrupo do sexo masculino (garantindo a comparabilidade dos grupos de intervenção e controlo assim formados em cada um dos subgrupos). Este processo permite ter um controlo adicional (além do garantido pela aleatorização) sobre variáveis de confusão mais importantes ou sobre variáveis que suspeitamos possam ser modificadoras do efeito da intervenção em estudo. Na generalidade, a estratificação só pode ser feita em função de um grupo restrito de variáveis (devido às limitações no tamanho da amostra). Sublinhe-se, no entanto, que no plano formal, só esta estratégia garante a possibilidade de interpretar os resultados das frequentemente usadas análises de subgrupos directamente como efeitos causais (ver abaixo)<sup>(16)</sup>.

A aleatorização permutada em blocos é um método utilizado para controlar as flutuações ao acaso dos tamanhos dos grupos a comparar<sup>(9)</sup>. Fazer uma aleatorização permutada em blocos implica não fazer atribuição aleatória da intervenção indivíduo a indivíduo, mas antes em grupos de indivíduos. Por exemplo, tipicamente uma aleatorização permutada em blocos de 10 elementos, num ECA para comparação de duas estratégias alternativas (com distribuição 1:1, isto é, com igual número de participantes nos dois grupos), é feita definindo à partida de forma aleatória, para cada conjunto sequencial de 10 participantes, qual a permutação escolhida com 5 elementos alocados a cada grupo (ex: ECCECE-EECC, em que o primeiro participante seria do grupo experimental, o segundo e terceiro do grupo de controlo, e assim sucessivamente). Este método garante

**Tabela 1:** Critérios de avaliação de ensaios clínicos aleatorizados relativamente a três aspectos fundamentais: (1) validade ou qualidade metodológica, (2) importância dos resultados e (3) aplicabilidade prática dos resultados.

---

## (A) Avaliação da validade ou qualidade metodológica:

---

1. Houve uma adequada atribuição aleatória das intervenções (aleatorização ou randomização)?
2. A lista de atribuições aleatórias foi ocultada aos profissionais de saúde ou investigadores responsáveis pelo recrutamento dos participantes e atribuição das intervenções?
3. Todos os participantes foram analisados nos grupos em que foram inicialmente alocados (análise segundo o princípio da intenção-de-tratar)?
4. Existiu ocultação (blinding ou masking) do grupo de intervenção a que cada participante pertence em todos os parceiros envolvidos na execução do estudo?
5. O seguimento (follow-up) dos participantes foi suficientemente longo e completo?
6. As variáveis de resultado clínico consideradas na análise são adequadas e adequadamente definidas?
7. As intervenções de comparação foram adequadamente seleccionadas?
8. Os grupos experimental e de controlo eram comparáveis, no início do estudo, relativamente a outras variáveis que possam eventualmente influenciar os resultados do mesmo?
9. São apresentados e adequadamente enquadrados e discutidos os resultados de análises de subgrupos?

---

## (B) Avaliação da importância científica e prática dos resultados

---

10. Qual é a magnitude e consequente importância científica, prática ou clínica do efeito da intervenção em estudo?
11. Qual é a precisão da estimativa de efeito da intervenção apresentada?

---

## (C) Avaliação da aplicabilidade prática dos resultados

---

12. Serão os resultados do estudo aplicáveis ao meu doente, tendo em conta as suas características específicas e individuais? Serão os resultados do estudo generalizáveis?
  13. No contexto onde me insiro, estarão as intervenções terapêuticas avaliadas disponíveis e serão estas aplicáveis na prática clínica?
  14. Quais são os potenciais benefícios e malefícios das intervenções terapêuticas no meu doente em particular?
  15. Quais são as opiniões, valores e expectativas do meu doente relativamente aos resultados clínicos esperados e à intervenção terapêutica proposta?
-

uma distribuição aleatória dos participantes pelos grupos, mas simultaneamente controla exactamente o tamanho de cada um dos grupos. Recomenda-se habitualmente a utilização de blocos de grande tamanho e, de preferência, com tamanhos de bloco a variar aleatoriamente.

A avaliação do artigo de Damoiseaux et al. [8], em análise neste caso, permite verificar que os autores implementaram, de facto, um processo aleatório de atribuição das intervenções, descrito na subsecção “Assignment and blinding” da secção “Methods”. Da descrição feita parece ter havido uma atribuição aleatória da intervenção, feita centralmente utilizando uma aplicação informática e aplicando um método de aleatorização permutada em blocos. Apesar de satisfatória, os autores poderiam ter sido um pouco mais detalhados na descrição do processo, do tamanho e tipo de blocos utilizados. A expressão usada no artigo “two block randomisation” não permite perceber claramente o que foi feito.

## **2) A lista de atribuições aleatórias foi ocultada aos profissionais de saúde ou investigadores responsáveis pelo recrutamento dos participantes e atribuição das intervenções?**

Um processo de atribuição aleatória da intervenção eficaz depende, na prática, de dois aspectos intrinsecamente ligados: a adequada geração de uma sequência imprevisível de atribuições e a ocultação dessa sequência aos responsáveis pelo recrutamento dos participantes, até a alocação da intervenção ter tido lugar<sup>(9,17-19)</sup>. Uma questão fundamental será garantir a imprevisibilidade da sequência de atribuições<sup>(20)</sup>. O método de alocação das intervenções deverá ser implementado de forma a garantir que seja impossível aos responsáveis pelo recrutamento dos participantes saber, prever ou manipular a intervenção a atribuir ao próximo participante recrutado. Esta característica é mais comumente conhecida pela sua designação em inglês – *allocation concealment*<sup>(17,18)</sup>. Existe evidência empírica que demonstra a importância deste critério, sendo que este impede que a acção consciente ou

inconsciente do clínico ou investigador responsável pelo recrutamento e alocação das intervenções possa perturbar a comparabilidade dos grupos criada pelo efeito da aleatorização<sup>(18,19,21)</sup>.

Infelizmente, a descrição que é feita nas publicações de ECA's relativamente a este ponto, não é geralmente muito satisfatória, sendo muitas vezes difícil julgar sobre a existência de um adequado *allocation concealment*<sup>(17-21)</sup>. Sempre que o processo de aleatorização e alocação das intervenções esteja a cargo de um centro de aleatorização, independente e fisicamente separado do local onde os participantes estão a ser recrutados, este critério estará a ser muito provavelmente cumprido. Outros métodos frequentemente usados para garantir este ponto são, por exemplo, a utilização de envelopes selados e sequencialmente numerados e o uso de embalagens incaracterísticas e sequencialmente numeradas<sup>(9,19)</sup>.

No artigo de Damoiseaux et al.<sup>(8)</sup> é possível verificar, na subsecção “Assignment and blinding” da secção “Methods”, que os autores do estudo utilizaram embalagens não identificáveis e uma lista de atribuições que era gerida por um centro independente (farmácia do hospital), garantindo desta forma uma adequada ocultação da lista de atribuições e do processo de alocação das intervenções.

## **3) Todos os participantes foram analisados nos grupos em que foram inicialmente alocados (análise segundo o princípio da intenção-de-tratar)?**

A análise de resultados num ECA deverá seguir, em primeira instância e de acordo com as recomendações nesta área<sup>(9,22,23)</sup>, o princípio de intenção-de-tratar (*intention-to-treat analysis* – ITT). Isto significa que a comparação dos grupos em avaliação, relativamente às variáveis de resultado clínico analisadas, deverá considerar todos os indivíduos alocados a cada um dos grupos, respeitando a atribuição aleatória da intervenção inicial, independentemente do que acontece depois dessa alocação; incluindo todos os participantes independentemente do tratamento de facto realizado, subsequente abandono do estudo ou alteração do pro-

toloco do mesmo<sup>(9,22-25)</sup>. Assim, deverão ser considerados na análise, e nos grupos inicialmente definidos, todos os participantes mesmo que estes: (1) tenham saído do estudo e não tenham seguido a intervenção que lhes foi atribuída (*drop-out*); (2) tenham, por qualquer razão, saltado do braço de controlo para o braço experimental (*drop-in*); (3) não tenham, total ou parcialmente, aderido ao tratamento atribuído (*noncompliance*); (4) tenham, por qualquer razão, deixado de ser seguidos no contexto do estudo (*lost to follow-up*); (5) tenham abandonado voluntariamente o estudo (*withdrawn*) ou (6) não tenham, por qualquer razão, seguido o protocolo do estudo (*protocol violations*).

A alternativa a uma análise segundo o princípio da intenção-de-tratar é uma análise em que são só considerados os participantes em função das intervenções de facto realizadas ou do protocolo de facto seguido (*on-treatment analysis* ou *per-protocol analysis*)<sup>(23,24)</sup>. Esta análise alternativa poderá complementar a análise segundo o princípio de intenção-de-tratar mas não deverá substituí-la<sup>(9)</sup>.

A análise ITT é mais adequada que as alternativas por três razões essenciais<sup>(9,23,25)</sup>. Em primeiro lugar, a análise ITT é o método mais apropriado para preservar a comparabilidade dos grupos que resulta da aleatorização inicialmente feita num ECA. A consideração unicamente dos participantes que são de facto submetidos às intervenções ou que completam o protocolo do estudo leva a uma avaliação enviesada do efeito da intervenção por romper a aleatorização inicial do estudo. Na generalidade, as violações do protocolo, as saídas do estudo ou a não-adesão estão associadas simultaneamente às intervenções (geralmente são mais frequentes no grupo experimental, devido por exemplo às reacções adversas) e às variáveis de resultado clínico (geralmente os doentes que saem ou abandonam têm pior prognóstico), logo, a sua exclusão leva à não comparabilidade dos grupos e ao enviesamento na avaliação do efeito da intervenção (geralmente sobrestimação). Em segundo lugar, a análise ITT responde à questão essencial e pragmática que tem, na generalidade, mais interesse para o clínico

– será que a decisão de utilizar determinada intervenção e a sua recomendação/prescrição estão associadas a um melhor resultado clínico? Na verdade, esta questão é mais relevante e diferente daquela que uma análise *per-protocol* ou *on-treatment* permite responder. Numa análise *per-protocol* a questão que é respondida é – será que a utilização de facto da intervenção, caso o indivíduo cumpra escrupulosamente o protocolo do estudo e as recomendações sobre a aplicação da intervenção, está associada a um melhor resultado clínico? Como já foi referido, a resposta que é possível dar neste tipo de análise será sempre potencialmente enviesada, para além de não ser geralmente a resposta mais útil para o clínico. Em terceiro lugar, uma análise ITT proporciona uma metodologia para o tratamento dos casos de quebras de protocolo, desistências e perdas de *follow-up* mais apropriada, mais segura e menos sujeita a potenciais manipulações<sup>(9,22)</sup>. Estas situações são frequentes e muitas vezes de difícil definição e tratamento, pelo que podem ser alvo fácil de manipulação. A análise ITT (se adequadamente implementada) obriga a um tratamento padronizado e não ambíguo destes casos<sup>(9,22)</sup>.

Assim, a apresentação de uma análise ITT é um critério importante e que está associado à qualidade metodológica global do ECA<sup>(22,23,25)</sup>. A verificação deste critério é, no entanto, por vezes difícil, dado que os artigos descrevem muitas vezes de forma incompleta a análise e o percurso de cada um dos participantes no ECA e utilizam de forma inadequada a expressão “análise segundo o princípio de intenção-de-tratar” (referindo-a mas não a aplicando). Na prática, a avaliação do tipo de análise feita implica a verificação dos participantes considerados na mesma, idealmente contemplando todos os indivíduos segundo as alocações iniciais. As únicas situações em que poderá ser lícito excluir participantes da análise são: (1) participantes que, por qualquer razão, são erradamente incluídos no estudo e que não cumprem os critérios de inclusão; (2) participantes que foram prematuramente aleatorizados e a quem lhes foi atribuída uma intervenção que, de facto, nunca chegaram a receber, desde que

a alocação ao grupo experimental não influencie a probabilidade de se receber a intervenção; (3) participantes para os quais não foi possível obter dados sobre as variáveis de resultado clínico analisadas (note-se que é obrigatório num ECA desenvolver todos os esforços para a obtenção dos dados relativos às variáveis de resultado em todos os participantes, especialmente no caso de abandonos ou perdas de seguimento), dado que naturalmente neste caso é impossível incluí-los na análise.

No artigo de Damoiseaux et al.<sup>(8)</sup> é possível verificar, no segundo parágrafo da subsecção “Sample size and data analysis” da secção “Methods”, que os autores referem ter utilizado uma análise segundo o princípio de intenção-de-tratar. A avaliação subsequente da secção de resultados permite confirmar que, de facto, a análise feita segue o princípio de intenção-de-tratar, dado que é possível verificar o que aconteceu com todos os participantes do ensaio no *flow-chart* apresentado nesta secção e é possível verificar, na tabela 2, que o cálculo das medidas de efeito consideraram todos os participantes (para os quais existem dados disponíveis) segundo as alocações iniciais.

#### **4) Existiu ocultação (*blinding* ou *masking*) do grupo de intervenção a que cada participante pertence em todos os parceiros envolvidos na execução do estudo?**

O termo ocultação (*blinding* ou *masking*) no contexto de um ECA refere-se à existência de desconhecimento do grupo de intervenção a que cada participante pertence por parte dos próprios participantes e dos investigadores e profissionais de saúde responsáveis pelo seu seguimento, pela recolha de dados, pela avaliação das variáveis de resultados ou pela análise dos resultados, de forma a garantir que este conhecimento não possa afectar o seu comportamento ao longo do ECA e enviesar os resultados do mesmo<sup>(26,27)</sup>. A ocultação é genericamente uma ferramenta metodológica muito importante, mas a sua relevância prática depende do contexto clínico, do tipo de intervenções e do tipo de variáveis de resultado clínico em análise<sup>(26,27)</sup>.

A ocultação dos participantes do estudo é fundamental para evitar a influência que o conhecimento da intervenção pode ter na manifestação e observação do resultado clínico; e visa evitar vieses relacionados com os efeitos não específicos da intervenção. A ocultação ao nível dos participantes e dos investigadores ou profissionais de saúde responsáveis pelo seguimento e recolha de dados visa evitar vieses relacionados com eventuais diferenças no seguimento, tratamento ou cuidados prestados (ou exigidos) aos participantes de diferentes grupos (vieses de desempenho – *performance bias*). A ocultação ao nível dos investigadores responsáveis pela avaliação das variáveis de resultado clínico é fundamental, pois impede que esta avaliação seja afectada consciente ou inconscientemente pelo conhecimento da intervenção e permite evitar vieses de informação, de observador ou de avaliação (*observer, ascertainment* ou *assessment bias*). A ocultação ao nível dos investigadores responsáveis pela análise dos dados é hoje também recomendada; e impede eventuais influências no tipo e profundidade da análise estatística efectuada<sup>(26-30)</sup>.

Existe abundante evidência empírica que demonstra a importância da ocultação nos vários níveis referidos, sendo que a não implementação ou inadequada descrição dos métodos de ocultação parecem estar associados a uma avaliação enviesada dos efeitos das intervenções, na generalidade sobrestimando-os<sup>(18,21,26,31)</sup>.

É frequente haver descrições incompletas dos métodos de ocultação utilizados e existe grande heterogeneidade no significado dado a expressões como “*double blinding*”. Por este motivo, recomenda-se que a descrição do processo de ocultação seja o mais completa e explícita possível e que sejam mencionados os vários níveis em que esta foi implementada e os métodos que foram utilizados [27-31]. A avaliação deste critério está dependente da verificação, para cada nível, do tipo e métodos de ocultação que foram implementados, e das garantias que estes dão sobre a eficácia da ocultação.

No artigo de Damoiseaux et al.<sup>(8)</sup> é possível verificar, na subsecção “Assignment and blinding” da secção “Methods”,

que os autores referem ter feito a ocultação ao nível das crianças participantes, dos seus pais e dos investigadores envolvidos no estudo. Apesar da referência, não são dadas explicações adicionais sobre os métodos de ocultação dos investigadores e sobre a ocultação aquando da avaliação das variáveis de resultado clínico. É referido que o placebo utilizado tinha uma cor e sabor idênticas ao antibiótico aplicado no grupo experimental, garantindo a ocultação das crianças e dos seus pais. É ainda descrito o facto de ter havido ocultação dos pais e investigadores como garantia de um tratamento igual dos participantes nos dois grupos ao longo do estudo.

### **5) O seguimento (*follow-up*) dos participantes foi suficientemente longo e completo?**

O seguimento dos participantes no contexto de um ECA é, muitas vezes, longo e complexo. É de todo o interesse, a bem da qualidade do estudo, que o seguimento dos participantes seja o mais completo possível, uma vez que as características e os resultados clínicos dos participantes perdidos são, na generalidade, diferentes daqueles que são completa e adequadamente seguidos<sup>(10,18,21)</sup>. É recomendada a implementação de métodos que visem maximizar a probabilidade de sucesso no seguimento dos participantes; e no caso de haver perdas, dever-se-á desenvolver esforços para obter o máximo de informação possível sobre o ocorrido e, de preferência, dados sobre as variáveis de resultado clínico. Não existem recomendações estritas sobre o que será um nível aceitável de perda de seguimento, no entanto, alguns autores usam a referência dos 80% de participantes com seguimento completo como o mínimo aceitável<sup>(10)</sup>.

A duração do seguimento é também um factor decisivo, já que esta deve garantir a possibilidade teórica e prática de observação, em todos os participantes, de todos os eventos clínicos relevantes relacionados com as variáveis de resultado importantes no contexto clínico em causa (ex: se estamos interessados na avaliação da mortalidade num ensaio sobre eficácia de anti-hipertensores em doentes de 50 anos, um período de se-

guimento de 1 mês ou 3 meses será claramente insuficiente)<sup>(10)</sup>. Simultaneamente, no entanto, o seguimento deverá ser o mais curto possível de forma a minimizar as perdas de seguimento e ser o menos intrusivo possível na vida dos participantes. O equilíbrio entre estes dois objectivos antagónicos é difícil mas deverá ser procurado da melhor forma.

A verificação da descrição feita sobre a magnitude e causas das perdas de seguimento e sobre a duração média do seguimento são, portanto, critérios fundamentais na avaliação da qualidade de ECA's<sup>(9-11)</sup>.

No artigo de Damoiseaux et al.<sup>(8)</sup> é possível verificar, no *flow-chart* e no texto da secção "Results", que os autores descrevem adequada e detalhadamente todos os casos em que existiu perda de seguimento e não foi possível obter informação sobre as variáveis de resultado clínico. Neste estudo existiu uma duração de seguimento adequada às variáveis de resultado clinicamente relevantes e as proporções de perda de seguimento foram muito satisfatórias (as proporções de participantes com seguimento completo foram de 98%, 91% e 88%, respectivamente para os 4 dias, 11 dias e 6 semanas de seguimento).

### **6) As variáveis de resultado clínico consideradas na análise são adequadas e adequadamente definidas?**

O desenho e execução de um ECA implica sempre a medição de variáveis de resultado clínico (*outcomes*) para comparação dos grupos e efectiva avaliação do efeito da intervenção em análise. As variáveis de resultado primárias (*primary outcomes*) são as variáveis de maior interesse no estudo, deverão ser em número limitado (idealmente só uma) e geralmente são utilizadas para o cálculo do tamanho da amostra. As variáveis de resultado secundárias são complementares relativamente às primárias, sendo importante que incluam, além de variáveis de eficácia, variáveis relativas à segurança das intervenções (ex: eventos ou reacções adversas)<sup>(9-11)</sup>.

A escolha de quais as variáveis de resultado clínico a considerar num ECA

deverá ter em conta as recomendações existentes na área clínica específica (ex: documentos de consenso de sociedades científicas internacionais) e, idealmente, estas deverão estar devidamente estudadas e validadas, de forma a garantir a qualidade das medições e a comparabilidade com outros estudos. As variáveis de resultado deverão ser descritas de forma detalhada e completa, indicando claramente quais as primárias e quais as secundárias. A descrição deverá incluir a sua definição, demonstração da sua relevância e indicação da metodologia seguida na sua medição. Aspectos importantes são, por exemplo, a indicação de quem fez a observação das variáveis; se houve ou não observadores múltiplos; se a observação foi com ou sem ocultação; se existiu mais que um ponto de avaliação ao longo do tempo (e se sim, qual o momento considerado na definição da variável de resultado) e o protocolo e tecnologia utilizada nas medições (se apropriado)<sup>(9-11)</sup>.

Dois problemas importantes relacionados com a definição de variáveis de resultado num ECA são: primeiro, a utilização de variáveis de resultado indirectamente relacionadas com as variáveis clinicamente mais importantes num determinado contexto (*surrogate outcomes* ou *endpoints*); segundo, a utilização de variáveis de resultado compostas, que resultam da agregação de vários eventos clinicamente relevantes mas distintos (*composite outcomes* ou *endpoints*).

Frequentemente num ECA a utilização das variáveis de resultado realmente relevantes do ponto de vista clínico (ex: mortalidade global, mortalidade cardiovascular, mortalidade devido a cancro, incapacidade a longo prazo, perda de acuidade visual ou de outras funções relevantes, etc.) podem implicar longos períodos de seguimento e um grande número de participantes. Por este motivo, e devido à pressão comercial e à necessidade de resultados rápidos e baratos, tem-se tornado cada vez mais frequente a utilização de variáveis de resultado que, apesar de não serem as de maior interesse clínico, se assume estarem directamente relacionadas com as variáveis clinicamente mais importantes (*surrogate outcomes*), permitindo fazer estudos mais curtos,

pequenos e baratos (ex: reduções de tamanho de tumores, diminuições de níveis de colesterol, diminuições de níveis de pressão arterial, contagens de células CD4+, etc.). Em muitos casos, este aspecto pode ser perfeitamente aceitável, no entanto, convém chamar a atenção para dois factos: primeiro, é fundamental estudar, sempre que possível, as variáveis que são na realidade clinicamente relevantes; segundo, frequentemente a assumption fundamental de uma relação estrita e linear entre os *surrogate outcomes* e as variáveis de resultado realmente importantes não é, por razões várias, verdadeira, implicando isto que não seja lícito concluir sobre os segundos na base de resultados sobre os primeiros<sup>(14,32)</sup>.

As variáveis de resultado compostas são frequentemente utilizadas e resultam da consideração conjunta de vários eventos potencialmente relevantes numa única variável. Por exemplo, é habitual em ensaios clínicos da área cardiovascular considerar uma variável composta de “eventos cardiovasculares relevantes” que resulta da junção de angina, enfarte do miocárdio não fatal e morte. Neste caso a variável “evento cardiovascular relevante” estará presente se qualquer um dos eventos que a compõem acontecer. A utilização de variáveis de resultado compostas é também uma estratégia que permite aumentar a eficiência estatística dos ensaios clínicos, no entanto, o seu uso reveste-se de vários problemas que devem ser considerados. Primeiro, uma variável composta junta frequentemente eventos com relevâncias muito diversas (ex: angina e morte); segundo, os eventos mais relevantes podem ser bastante mais raros que os restantes (ex: um efeito detectado poderá reflectir maioritariamente reduções na angina e não na mortalidade); terceiro, o efeito da intervenção pode ser diferente nos vários componentes da variável composta (ex: uma intervenção poderá ter um efeito na redução da angina mas não na mortalidade). Assim, se os eventos clinicamente mais relevantes são raros ou se há suspeita da existência de efeitos diferentes nos vários componentes, não será aconselhável olhar só para as variáveis compostas, mas antes verificar o efeito da intervenção

nos vários componentes isolados, ainda que isto possa implicar algumas perdas na precisão das estimativas dos efeitos (relacionada com a raridade dos eventos mais relevantes)<sup>(33)</sup>.

No artigo de Damoiseaux et al.<sup>(8)</sup> é possível verificar, na subsecção “Outcome measures” da secção “Methods”, que existe uma descrição detalhada e completa das variáveis de resultado primárias e secundárias consideradas no estudo. Foram utilizadas as variáveis de resultado clínico geralmente encontradas em estudos desta natureza, são descritos os métodos utilizados na sua medição e, no caso do diagnóstico de derrame no ouvido médio às 6 semanas, apresentadas referências que suportam a validade do método utilizado. A variável de resultado primária aplicada neste estudo foi a persistência de sintomas ao quarto dia, avaliada pelo médico assistente e definida como a persistência de otalgia, febre ( $\geq 38^{\circ}\text{C}$ ), choro ou irritabilidade, este é um exemplo de uma variável de resultado composta, dado que implica a presença, aos 4 dias, de qualquer um dos 4 eventos componentes. Embora clinicamente muito útil, esta variável permite, eventualmente, classificar da mesma forma uma criança unicamente com persistência de otalgia (não muito fácil de avaliar nesta idade) e uma criança com persistência de febre e irritabilidade aos 4 dias. Este exemplo extremo ilustra o problema do uso das variáveis compostas, no entanto neste caso a sua utilização revela-se útil e necessária. A variável de falência terapêutica aos 11 dias utilizada no estudo é também um exemplo de uma variável composta.

### **7) As intervenções de comparação foram adequadamente seleccionadas?**

Várias revisões sistemáticas têm demonstrado que ensaios clínicos realizados pela indústria farmacêutica tendem a apresentar efeitos de intervenções mais positivos que os ensaios realizados por entidades sem fins lucrativos<sup>(34,35)</sup>. Uma das aparentes razões deste fenómeno está relacionada com a escolha apropriada das intervenções de comparação<sup>(33,35,36)</sup>. Apesar de ensaios clínicos

que utilizam o placebo como comparador serem muito importantes no plano científico e regulamentar, clinicamente a comparação mais relevante será, na maioria das vezes, com as outras intervenções já habitualmente utilizadas no contexto clínico em causa; sendo do interesse do clínico a vantagem adicional que a nova intervenção poderá ter relativamente às armas terapêuticas já disponíveis e não relativamente ao placebo. Neste contexto, será importante também considerar a escolha de comparadores adequados e justos, por exemplo, em doses adequadas (ex: será muito fácil demonstrar maior eficácia do que uma intervenção alternativa em doses sub-terapêuticas) e utilizando vias de administração apropriadas (ex: se a intervenção de comparação não tem boa absorção oral, a sua utilização por via oral poderá levar-nos a observar um efeito inexistente)<sup>(33)</sup>.

Destaca-se, no entanto, que quando a comparação não pode ser feita com outros tratamentos activos, a utilização do placebo é de extrema importância, uma vez que permite controlar o efeito não específico relacionado unicamente com o facto de os indivíduos estarem a ser submetidos a uma intervenção (mesmo que esta não tenha um efeito activo demonstrado no contexto clínico em causa) – efeito placebo<sup>(37,40)</sup>. Daí a sua importância científica e regulamentar. Um outro efeito não específico comum neste contexto é, por exemplo, o chamado efeito de Hawthorne, que resulta unicamente do facto de os indivíduos estarem a ser seguidos num estudo<sup>(37,41)</sup>.

Na avaliação deste critério dever-se-á ter em atenção a escolha que foi feita da intervenção de comparação. Dever-se-á avaliar o que será mais importante no contexto clínico em causa, se uma comparação com placebo ou uma comparação com um tratamento habitual já disponível. Quando aplicável, dever-se-á ter extremo cuidado na análise da forma farmacêutica, dose e método de administração da intervenção de comparação.

No artigo de Damoiseaux et al.<sup>(8)</sup> é possível verificar, na subsecção “Intervention” da secção “Methods”, que as intervenções avaliadas neste caso são o tratamento habitual mais a utilização de

placebo (representativo de uma alternativa de observação expectante e tratamento habitual, sem inclusão de antibiótico) versus o tratamento habitual mais a utilização de antibiótico (amoxicilina 40 mg/Kg/dia). No artigo é possível verificar uma cuidadosa e completa descrição das intervenções comparadas e verifica-se também que as doses e vias de administração do antibiótico são as esperadas neste contexto.

#### **8) Os grupos experimental e de controlo eram comparáveis, no início do estudo, relativamente a outras variáveis que possam eventualmente influenciar os resultados do mesmo?**

Se o processo de atribuição aleatória da intervenção foi eficaz, espera-se que entre os grupos haja uma distribuição homogénea das características mais relevantes na determinação das variáveis de resultado em estudo. Em geral, os artigos que reportam resultados de ECA's apresentam uma tabela inicial em que são comparados os grupos experimental e de controlo relativamente a um conjunto de variáveis. A avaliação desta tabela e das diferenças encontradas entre os grupos permitirá ter uma noção sobre a eficácia do processo de aleatorização e sobre a necessidade de aplicar ajustamentos estatísticos relacionados com as diferenças detectadas (potenciais variáveis de confusão)<sup>(9-11)</sup>.

No artigo de Damoiseaux et al.<sup>(8)</sup> é possível verificar, na secção "Results", que os autores descrevem, na tabela 1, as características e variáveis clínicas mais importantes neste contexto, permitindo avaliar a comparabilidade dos grupos. Nesta tabela e no texto da secção de resultados é possível observar a descrição de diferenças significativas entre os grupos à partida, relativamente às variáveis antecedentes de recorrência de OMA, frequência de infantário e hábitos tabágicos dos pais. Na subsecção "Sample size and data analysis" da secção "Methods" e na secção "Results", é possível verificar que os autores descrevem, correctamente, ter complementado a sua análise estatística através da aplicação de uma técnica de modelação

que permite ajustar a estimativa do efeito da intervenção para estes desequilíbrios encontrados nestas variáveis (modelo de regressão logística múltipla). A aplicação desta técnica de modelação permitiu verificar que o ajustamento para essas diferenças não alterou minimamente a estimativa de efeito apresentada.

#### **9) São apresentados e adequadamente enquadrados e discutidos os resultados de análises de subgrupos?**

É muito frequente, especialmente em ensaios com resultados negativos (ausência de efeito da intervenção em análise), encontrar reportados resultados sobre efeitos significativos (portanto aparentes resultados positivos) em determinados subgrupos de doentes. A apresentação e a interpretação de resultados de subgrupos deverão ser feitas com extremo cuidado, pois a comparabilidade dos grupos, que é garantida pelo processo de aleatorização, só é aplicável aos participantes na sua globalidade e não aos subgrupos<sup>(42-44)</sup>. Portanto, na análise de subgrupos a apresentação de resultados estará sempre potencialmente afectada pelo enviesamento relacionado com a quebra da aleatorização inicial<sup>(9-11,42-44)</sup>. As únicas excepções, como referido anteriormente, são os casos em que foi utilizada uma aleatorização estratificada em função da variável que define os subgrupos, estando nesses casos a comparabilidade dos grupos salvaguardada também nos subgrupos<sup>(16)</sup>. A análise de subgrupos está também associada ao aumento da probabilidade de erros devido a fenómenos do acaso, uma vez que, naturalmente, se o investigador olhar para um número suficientemente grande de subgrupos irá, muito provavelmente, encontrar alguns efeitos estatisticamente significativos que são, no entanto, meros achados devidos ao acaso e não efeitos reais<sup>(43)</sup>.

Desta forma, os resultados de análises de subgrupos deverão ser avaliados sempre com extremo cuidado<sup>(44)</sup>. Recomenda-se que estas só sejam consideradas hipóteses credíveis quando cumulativamente foram verificadas as seguintes condições: (1) o efeito específico no subgrupo faz sentido do ponto

de vista clínico e biológico; (2) o efeito no subgrupo é clínico e estatisticamente significativo; (3) o efeito no subgrupo foi colocado como hipótese inicialmente, sendo referido o seu estudo no protocolo do ECA (não sendo, portanto, produto de análises exploratórias dos dados após a realização do estudo); (4) o efeito no subgrupo foi já verificado por outros estudos independentes<sup>(10)</sup>.

No artigo de Damoiseaux et al.<sup>(8)</sup> não são apresentados quaisquer resultados relativos a análise de subgrupos.

#### **(B) Avaliação da importância científica e prática dos resultados**

A avaliação da importância dos resultados passa pela adequada interpretação e contextualização das medidas e estatísticas aplicadas na apresentação dos mesmos. Genericamente, a avaliação da importância dos resultados deverá considerar dois aspectos essenciais<sup>(3,45)</sup>:

#### **10) Qual é a magnitude e consequente importância científica, prática ou clínica do efeito da intervenção em estudo?**

Este ponto implica a avaliação das medidas e magnitude dos efeitos apresentados, a sua adequada interpretação e a sua devida contextualização clínica.

A situação mais habitual em ECA's na área biomédica é a consideração de variáveis de resultado clínico categóricas, geralmente dicotómicas e relacionadas com a ocorrência ou não de um determinado evento clinicamente relevante (ex: morte, enfarte agudo do miocárdio, resolução clínica de sintomas, etc.), ao longo do seguimento prospectivo dos participantes do estudo. Neste contexto, as medidas de efeito mais comumente encontradas, hoje em dia, na apresentação de resultados de ECA's são o risco relativo (RR), a redução absoluta de risco (RAR), a redução relativa de risco (RRR) e o número-necessário-para-tratar (NNT)<sup>(10)</sup>.

Central ao cálculo destas medidas é o conceito de risco de desenvolvimento de em determinado evento. O risco é definido como a probabilidade de desenvolvimento de um evento ao longo de um determinado tempo de seguimento, este está intimamente associado aos concei-

tos de incidência cumulativa do evento ou da taxa de incidência do mesmo, dependendo do desenho do estudo realizado (a discussão da diferença entre estas duas medidas está fora do âmbito deste artigo). Considere-se, por exemplo, um ECA em doentes internados em unidade de cuidados intensivos (UCI) e em que foram seguidos um grupo experimental submetido a uma nova intervenção, que visa reduzir a mortalidade destes doentes, e um grupo de controlo submetido a um placebo. Imaginemos que cada grupo era constituído por 100 indivíduos que foram seguidos até ao fim do internamento hospitalar, tendo-se verificado que 20 do grupo controlo e 10 do grupo experimental morreram durante esse período. Neste exemplo simples temos duas medidas de risco fundamentais:

Risco do evento no grupo de controlo (REC) – Definido como a probabilidade de ocorrência do evento (morte) durante o período de seguimento (neste caso considera-se o período de internamento hospitalar), podendo então afirmar-se que a mortalidade hospitalar (risco de morte) no grupo de controlo é estimada através do cálculo da proporção de participantes deste grupo que morreram durante esse período ( $REC = 20/100 = 0,2 = 20\%$ ).

Risco do evento no grupo experimental (REE) – Definido da mesma forma, poderá afirmar-se que a mortalidade hospitalar no grupo experimental é estimada pela proporção de participantes deste grupo que morreram durante esse período ( $REE = 10/100 = 0,1 = 10\%$ ).

A partir destas medidas de risco será possível calcular as medidas de efeito (também designadas medidas de associação) atrás referidas:

O risco relativo (RR) – É uma medida relativa de efeito, estimada a partir da razão entre o risco no grupo experimental e o risco no grupo de controlo ( $RR = REE/REC = 0,1/0,2 = 0,5$ ), dando uma noção relativa da frequência do evento no grupo experimental em comparação com o grupo de controlo. Neste caso, um RR de 0,5 significa que existem 1 evento no grupo experimental para cada 2 no grupo de controlo (1/2), ou seja, o evento é 2 vezes mais frequente no grupo de controlo relativamente ao experimental.

A redução absoluta de risco (RAR) – É uma medida absoluta de efeito, estimada a partir do valor absoluto da diferença entre o risco no grupo de controlo e o risco no grupo experimental ( $RAR = REC - REE = 0,2 - 0,1 = 0,1 = 10\%$ ). Esta medida assume uma redução do risco do evento no grupo de controlo e é interpretada como a redução absoluta do risco do evento associada à intervenção, neste caso, poder-se-á afirmar que o efeito da intervenção implica uma redução do risco de morte de 10%.

A redução relativa de risco (RRR) – É uma medida de impacto potencial (essencialmente não é uma medida de efeito ou de associação), estimada a partir da razão entre a redução absoluta de risco e o risco no grupo de controlo ( $RRR = RAR/REC = 0,1/0,2 = 0,5 = 50\%$ ). Esta medida estima o impacto da intervenção sobre o risco basal do evento na população a que os participantes pertencem e é interpretada como a proporção desse risco basal que poderá ser potencialmente reduzida pela implementação da intervenção. Neste caso, temos como estimativa do risco basal da população o risco no grupo de controlo (estimativa do risco do evento quando a intervenção não está presente) e assumimos que a intervenção provoca uma redução desse risco basal que é estimada pela RAR. Assim, a proporção de risco basal que é aparentemente reduzida pela intervenção é dada pela razão entre o RAR e o REC, isto é, 50%. Portanto a intervenção parece potencialmente reduzir em 50% o risco do evento.

O número-necessário-para-tratar (NNT) – É uma medida absoluta, estimada a partir do inverso aritmético da redução absoluta de risco ( $NNT = 1/RAR = 1/0,1 = 10$ ). Esta medida estima o número de indivíduos que seria necessário tratar com a intervenção em avaliação de forma a prevenir um único evento (morte). Neste caso, seria necessário tratar 10 indivíduos com a intervenção de forma a prevenir uma morte. Esta é uma medida que tem vindo a ser crescentemente utilizada dada a sua interpretabilidade, a utilidade clínica do seu resultado, e por permitir comparar intervenções de uma forma mais fácil e adequada.

A compreensão da utilidade clínica do NNT e das vantagens e desvantagens

das medidas relativas atrás mencionadas, poderá ser melhor ilustrada através de um segundo exemplo trivial em que o ECA atrás descrito foi repetido, exactamente da mesma forma, para uma outra intervenção, mas os resultados desta vez indicam a existência de uma morte no grupo experimental e 2 mortes no grupo de controlo. Neste segundo exemplo, teremos que o  $REC = 0,02 = 2\%$ ; o  $REE = 0,01 = 1\%$ ; o  $RR = 0,01/0,02 = 0,5$ ; a  $RAR = 0,02 - 0,01 = 0,01 = 1\%$ ; a  $RRR = 0,01/0,02 = 50\%$  e o  $NNT = 1/0,01 = 100$ . Este exemplo ilustra que, apesar de esta intervenção ser muito diferente da anterior, por condicionar uma redução do risco de morte de apenas 1%, portanto, dez vezes inferior à anterior, tem medidas de efeito relativas com o mesmo valor. Para as duas intervenções observa-se um  $RR = 0,5$  e uma  $RRR = 50\%$ . Um investigador interessado em apresentar evidência da vantagem da intervenção poderá apresentar selectivamente as medidas relativas e não as restantes, eventualmente induzindo em erro o leitor. Repare-se que, do ponto de vista clínico, será mais fácil perceber as diferenças entre as duas intervenções ao observar e interpretar a  $RAR = 1\%$  e o  $NNT = 100$ . Percebemos claramente que a primeira intervenção é clinicamente mais importante pois condiciona uma redução absoluta do risco de morte 10 vezes maior e é necessário tratar um número muito menor de doentes para prevenir uma morte. Idealmente a apresentação de medidas absolutas de efeito deverá sempre acompanhar a apresentação das medidas relativas.

Atente-se ao facto de medidas como a RAR, a RRR e o NNT assumirem intrinsecamente a redução de risco de um evento lesivo ou adverso no grupo experimental relativamente ao controlo. Situações em que sejam considerados aumentos de risco de um evento lesivo ou adverso no grupo experimental relativamente ao controlo poderão levar ao cálculo análogo de medidas como o aumento absoluto de risco ( $AAR = REE - REC$ ), o aumento relativo de risco ( $ARR = AAR/REE$ ) e o número-necessário-para-lesar ( $NNL = 1/AAR$ ; do inglês *NNH – number-needed-to-harm*).

Obviamente, existem muitas situações em que num ECA as variáveis

de resultado clínico consideradas não são categóricas dicotômicas, mas antes quantitativas contínuas ou discretas ou categóricas ordinais. Nesses casos as medidas atrás referidas não são directamente aplicáveis, no entanto existem medidas específicas para estas outras escalas (ex: diferença entre as médias) que permitem a comparação dos grupos de uma forma análoga; e por vezes há lugar à categorização dessas variáveis permitindo a aplicação indirecta das medidas acima referidas.

No artigo de Damoiseaux et al. [8] é possível verificar, na tabela 2 da secção “Results”, que relativamente à variável de resultado primária – persistência de sintomas aos 4 dias – é descrita uma diferença estatisticamente significativa entre os grupos com uma RAR de 13% (risco no grupo placebo de 72% e no grupo de amoxicilina de 59%), correspondendo a um NNT de 8 crianças. Adicionalmente, foram também encontradas diferenças estatisticamente significativas relativamente à duração da febre (mediana de 2 dias no grupo da amoxicilina e mediana de 3 dias no grupo placebo) e ao consumo de analgésicos durante os primeiros 10 dias (média de 2,3 doses no grupo da amoxicilina e média de 4,1 doses no grupo placebo). Estas diferenças poderão ser vistas eventualmente como relevantes, no entanto, é fundamental sublinhar que os autores não encontraram diferenças significativas entre os grupos estudados relativamente ao exame objectivo da membrana timpânica aos 4 dias, proporção de falência terapêutica aos 11 dias, a duração de sintomas como a dor e o choro e na presença de derrame no ouvido médio, recorrências, uso de antibióticos e referências a especialistas às 6 semanas.

#### **11) Qual é a precisão da estimativa de efeito da intervenção apresentada?**

Este ponto implica a verificação das medidas de precisão das estimativas dos efeitos apresentados, geralmente na forma de intervalos de confiança.

As medidas de efeito apresentadas num ECA 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 efeito resultante do estudo de uma amostra de 1000 participantes será mais precisa (menos sensível aos erros aleatórios relacionados com o processo de amostragem) que uma estimativa proveniente de um estudo com 50 participantes. O tamanho da amostra é um dos mais importantes determinantes da precisão da estimativa de efeito, no entanto, a frequência do evento nos grupos, a variância da variável de resultado considerada e o nível de confiança assumido são factores também relevantes. A forma mais comum de reportar a precisão da estimativa de um efeito é através da apresentação do seu intervalo de confiança, geralmente de 95% (IC 95%). A definição formal mais correcta diz-nos que os IC 95% são intervalos que, construídos na base da metodologia de inferência estatística aplicável, nos dão a garantia que, caso a experiência em causa fosse repetida muitas vezes, 95% dos intervalos assim construídos conteriam dentro de si o verdadeiro valor do parâmetro populacional que pretendemos estimar. Numa simplificação deste conceito, muitas vezes usada por ser mais útil (mas formalmente menos correcta), pode-se pensar nos IC 95% como intervalos onde podemos com 95% de certeza encontrar o valor do parâmetro populacional.

Assim, é fundamental perceber qual a qualidade e precisão das estimativas de efeito apresentadas num ECA, através da avaliação dos seus IC 95%, já que disso dependerá a interpretação e potencial aplicação prática que lhes podemos dar.

No artigo de Damoiseaux et al. [8] é possível verificar, na tabela 2 da secção “Results”, que relativamente à variável de resultado primária – persistência de sintomas aos 4 dias – é descrita uma diferença estatisticamente significativa entre os grupos com uma RAR de 13% (com maior risco de persistência de sintomas no grupo placebo), sendo correctamente apresentado também o IC 95% para esta estimativa [1% – 25%]. Deverá ter-se em atenção que os dados apresentados são compatíveis tanto com resultados bastan-

te modestos de 1% como com resultados de 25%, pelo que a estimativa não é particularmente precisa. Na tabela 2 do artigo são igualmente reportados os IC 95% para as outras estimativas apresentadas.

#### **(C) Avaliação da aplicabilidade prática da evidência**

A última questão a ser abordada no processo de avaliação crítica da evidência será a questão da aplicabilidade prática da evidência científica encontrada. O objectivo do profissional de saúde será, na maior parte dos casos, a eventual aplicação da evidência científica aos seus problemas clínicos e aos seus doentes, logo, a avaliação da aplicabilidade prática é uma questão fundamental neste contexto<sup>(3,13,45)</sup>. Esta avaliação deverá, idealmente, contemplar três aspectos essenciais – questões biológicas, questões socioeconómicas e questões epidemiológicas<sup>(3)</sup>. O objectivo fundamental será o adequado enquadramento da evidência científica, dadas as características específicas do profissional de saúde, do contexto da sua prática clínica e das características específicas do seu doente. Os critérios fundamentais para avaliação da aplicabilidade prática de evidência científica são os seguintes:

#### **12) Serão os resultados do estudo aplicáveis ao meu doente, tendo em conta as suas características específicas e individuais? Serão os resultados do estudo generalizáveis?**

A avaliação da aplicabilidade dos resultados de um ECA ao meu doente passa, em primeira instância, pela análise crítica dos critérios de selecção (inclusão e exclusão) dos participantes aplicados no estudo. A utilização de critérios estritos de selecção de participantes num ECA é uma ferramenta metodológica importante, pois permite maximizar a validade interna do estudo; proporcionar condições experimentais mais controladas; diminuir perdas de seguimento e desistências; excluir subgrupos particularmente vulneráveis ou com riscos particulares (ex: grávidas, crianças, etc.). No entanto, o reverso da medalha é que quanto mais estritos forem os critérios de selecção, menor será a va-

lidade externa do estudo e, portanto, menor será a capacidade de generalização dos resultados do estudo para a população em causa, para outras populações e, em última análise, para o meu doente. A avaliação crítica dos critérios de selecção de um ECA implica a comparação das características do meu doente (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 àquele doente em particular. Dever-se-á ter em conta, no entanto, que a maior parte das vezes as diferenças não serão, de facto, tão grandes que justifiquem ignorar por completo os resultados do ECA, e na ausência de melhor evidência, estes poderão constituir o melhor ponto de partida disponível<sup>(10)</sup>.

Ao considerar a comparação entre as características individuais do meu doente e as características dos participantes no estudo dever-se-á fazer uma chamada de atenção especial para a utilização de resultados provenientes de análises de subgrupos (uma tentação grande neste contexto). Lembra-se que as análises de subgrupos num ECA devem ser vistas sempre com grande cuidado e a sua consideração deverá ter sempre em conta os critérios que atrás foram referidos.

No artigo de Damoiseaux et al.<sup>(8)</sup> é possível verificar, na subsecção "Study population" da secção "Methods", que no estudo foram incluídas 240 crianças entre os 6 e os 24 meses de idade, recrutadas no contexto dos cuidados primários de saúde numa região da Holanda, com um quadro de OMA (segundo a definição padrão usada na Holanda e aceite mundialmente); tendo sido considerados como critérios de exclusão o uso de antibióticos nas 4 semanas precedentes, alergia comprovada à amoxicilina, imunodepressão, anomalias crânio-faciais ou Síndrome de Down. Neste caso, parece sensato concluir que os resultados deste estudo são passíveis de serem generalizados, por exemplo, para a população de crianças portuguesas, sendo muito provavelmente aplicáveis a uma qualquer criança portuguesa nesse grupo etário com um quadro de OMA sem outras complicações ou factores de risco.

### **13) No contexto onde me insiro, estarão as intervenções terapêuticas avaliadas disponíveis e serão estas aplicáveis na prática clínica?**

Um ponto adicional na avaliação da aplicabilidade dos resultados provenientes de um ECA será a verificação de questões práticas relacionadas com a aplicação da intervenção em causa, nomeadamente, a disponibilidade da intervenção no meu contexto clínico específico; a disponibilidade dos meios necessários para a administração e monitorização da intervenção e o eventual seguimento dos doentes; a disponibilidade do sistema de saúde para financiar a intervenção e a sua custo-efectividade<sup>(10)</sup>.

No artigo de Damoiseaux et al. são comparados o tratamento com amoxicilina versus placebo, em crianças com OMA e menos de dois anos de idade. Estas duas alternativas são perfeitamente aplicáveis e já em uso em qualquer contexto habitual. Neste caso, deverão adicionalmente ser ponderados os riscos associados à eventual utilização indevida de antibióticos (reações adversas desnecessárias, o problema das resistências aos antibióticos e os custos desnecessários).

### **14) Quais são os potenciais benefícios e malefícios das intervenções terapêuticas no meu doente em particular?**

É fundamental ter em conta na avaliação da evidência e na tomada de decisão tanto os dados relativos à eficácia da intervenção quanto os dados relativos aos potenciais malefícios associados à mesma (risco de reacções ou eventos adversos).

Caso tenhamos concluído que a intervenção em causa é aplicável e a evidência da sua eficácia é válida e clinicamente importante, o passo seguinte na decisão será a de considerar as características específicas do meu doente para estimar os seus potenciais benefícios e malefícios. Para isto dever-se-á, de forma mais ou menos quantitativa<sup>(10)</sup>, ajustar as estimativas de efeito da intervenção e as estimativas de probabilidade de reacções ou eventos adversos às características clínicas particulares do doente.

No artigo de Damoiseaux et al.<sup>(8)</sup> são comparados o tratamento com antibiótico versus o placebo (tido como representativo da alternativa de observação expectante sem tratamento antibiótico), em crianças com menos de dois anos com OMA, tendo-se verificado que o benefício do antibiótico é uma redução significativa modesta da persistência de sintomas aos 4 dias, com uma RAR de 13% e um NNT de 8, e uma redução de um dia na mediana de duração da febre. Não foi possível demonstrar diferenças significativas entre os grupos no exame objectivo da membrana timpânica aos 4 dias; na proporção de falência terapêutica aos 11 dias e na presença de derrame no ouvido médio, recorrências, uso de antibióticos e referências a especialistas às 6 semanas. Adicionalmente, neste artigo são reportadas reacções adversas, tendo-se verificado que o potencial malefício do antibiótico passava por um aumento do risco de diarreia aos 4 dias, com um AAR de 7% e um NNL de 14.

### **15) Quais são as opiniões, valores e expectativas do meu doente relativamente aos resultados clínicos esperados e à intervenção terapêutica proposta?**

Finalmente, a decisão sobre a aplicabilidade dos resultados do ECA ao meu doente deverá ter em conta o equilíbrio entre benefícios e malefícios da intervenção no seu caso 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 os potenciais malefícios da intervenção como factores de ponderação na comparação entre estes. Para isto, dever-se-ão utilizar métodos mais ou menos quantitativos e mais ou menos explícitos de elicitación e integração das valorizações dos doentes<sup>(10)</sup>. Por exemplo, se um doente atribui um peso grande à valorização das potenciais reacções adversas que podem advir da aplicação da intervenção, isto deverá ser tido em consideração na decisão, mesmo que esta tenha um benefício claro demonstrado. Num caso extremo, e mesmo que a intervenção possa ter um claro benefício, o doente poderá optar por não aderir à intervenção por causa

das suas potenciais reacções adversas. Se isto acontecer, é preferível que seja no momento da decisão, com o conhecimento e apoio do médico e integrando-se no próprio processo de decisão, em vez de ser feito depois da decisão do clínico e sem o seu conhecimento.

No artigo de Damoiseaux et al.<sup>(8)</sup> são apresentados resultados que demonstram a existência de potenciais benefícios e malefícios com o uso de antibióticos no tratamento de crianças com menos de dois anos com OMA. Tendo em conta estes resultados, o cenário que deverá ser explicado ao doente, ou neste caso aos seus pais, é que a utilização de antibióticos no tratamento da OMA deverá ser ponderada pensando, por um lado, na valorização de uma modesta redução do risco de persistência de sintomas aos 4 dias de 13% e NNT de 8 crianças e, por outro lado, na valorização de um aumento do risco de diarreia aos 4 dias de 7% e NNT de 14 crianças. Acrescentando-se que as avaliações aos 11 dias e às 6 semanas não demonstraram quaisquer vantagens significativas do uso de antibióticos. É perfeitamente plausível pensar que na maioria dos casos a valorização dos potenciais benefícios e dos potenciais malefícios, neste caso, sejam muito próximos. Se aceitarmos este pressuposto, torna-se ainda menos importante o já por si pequeno benefício que o uso de antibióticos parece ter neste caso.

## CONCLUSÃO

Os ensaios clínicos aleatorizados são o tipo de estudos primários considerados como padrão metodológico na resposta a questões sobre a eficácia de intervenções terapêuticas, devido ao seu carácter controlado e à aplicação da aleatorização. Ao longo deste artigo foi possível verificar que a qualidade destes está dependente do controlo de um vasto conjunto de importantes 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 de ECA's e da importância e aplicabilidade dos seus resultados foi discutida e ilustrada através de um exemplo prático.

O ensaio clínico de Damoiseaux et al.<sup>(8)</sup>, sobre a eficácia dos antibióticos no tratamento de OMA, em crianças com menos de dois anos, foi apresentado e discutido ao longo deste artigo. Verificou-se que este estudo apresenta uma excelente qualidade metodológica, sendo os únicos aspectos a apontar a necessidade de uma descrição um pouco mais clara e completa dos métodos de aleatorização e ocultação.

Voltando ao caso específico do pediatra assistente do Manuel, que é uma criança de 18 meses de idade com o diagnóstico de OMA sem complicações ou factores de risco, a decisão a tomar será a de prescrever ou não antibióticos, tendo em atenção os resultados do estudo discutido. Neste contexto, dever-se-á reflectir sobre quais são os objectivos clinicamente mais relevantes no tratamento da OMA numa criança e qual a sua ordem de prioridade. Será eventualmente consensual que os objectivos do tratamento são o alívio sintomático (a curto prazo) e a prevenção de complicações e sequelas que possam advir de um episódio de OMA (a curto, médio e longo prazo).

O estudo de Damoiseaux et al. demonstra um efeito significativo moderado dos antibióticos no alívio sintomático (RAR de 13% e NNT de 8), que tem que ser equacionado com um risco de reacções adversas – diarreia – de magnitude comparável (AAR de 7% e NNT de 14), e uma ausência de efeito significativo no que às variáveis relacionadas com a prevenção de complicações ou sequelas diz respeito. Na verdade, os mesmos autores publicaram recentemente um outro estudo<sup>(46)</sup>, onde apresentam os resultados do seguimento de três anos das crianças incluídas no ECA analisado, tendo concluído que, durante estes três anos de seguimento, não foram observadas diferenças significativas entre as crianças nos dois grupos (antibiótico versus placebo) na frequência de referências a especialistas nos cuidados secundários (pediatras ou otorrinolaringologistas) e na frequência de cirurgias do foro de otorrinolaringologia; mas encontraram um aumento estatisticamente significativo de 20% (IC 95% [5% – 35%]) no risco de recorrências de OMA no grupo submetido a antibiótico. Este in-

teressante resultado leva os autores a colocar a hipótese de este poder dever-se ao efeito do uso da antibioterapia no aumento da frequência de estirpes resistentes.

Embora os resultados do estudo de Damoiseaux et al.<sup>(8)</sup> sejam adequados e importantes, será sensato verificar os resultados de outros ensaios clínicos para verificar se estes são confirmados. A forma mais eficiente de fazer isto é através da procura e avaliação crítica de revisões sistemáticas e meta-análises que tenham respondido à mesma questão de investigação. Apesar da análise crítica deste tipo de estudos estar fora do âmbito deste artigo (tendo já sido focada no número anterior desta série), verifica-se a existência de pelo menos quatro revisões sistemáticas e meta-análises neste âmbito já publicadas<sup>(47-50)</sup>. As conclusões apontadas nestes estudos são semelhantes às já descritas, ou seja, o antibiótico está associado a uma moderada redução na persistência de sintomas e a um aumento de magnitude comparável no risco de reacções adversas como diarreia, não sendo melhor que o placebo ou a observação expectante na redução falências terapêuticas ou de complicações ou sequelas a curto ou médio prazo. Os estudos incluídos nestas revisões sistemáticas incluem maioritariamente participantes entre os 6 meses e os 12 anos de idade, no entanto, não é apresentada evidência clara que demonstre diferenças destes resultados específicas para o grupo etário abaixo dos 2 anos de idade.

Assim, na base da evidência científica apresentada e discutida, a decisão sobre a utilização de antibióticos no tratamento de OMA's, em crianças com menos de 2 anos de idade, deverá ter em conta que: (1) estes têm um benefício moderado na melhoria sintomática, (2) têm um risco de reacções adversas de nível semelhante, (3) não diminuem o risco de falências terapêuticas, complicações ou sequelas a curto e médio prazo, (4) têm o potencial de agravar o problema das resistências aos antibióticos e (5) têm custos para o sistema de saúde associados.

Voltando finalmente ao cenário clínico apresentado, o pediatra assistente do Manuel, depois de uma conversa es-

clarecedora com a mãe e num processo de tomada de decisão partilhada, optou por não prescrever antibióticos, tendo recomendado a utilização de medicação analgésica e antipirética, hidratação adequada e uma reavaliação no caso de persistência ou agravamento dos sintomas. Embora ainda com algumas apreensões, a mãe do Manuel decidiu que seria melhor, atendendo ao bem-estar futuro do seu filho, tentar seguir esta estratégia.

### CRITICAL APPRAISAL AND PRACTICAL IMPLEMENTATION OF RANDOMISED CONTROLLED TRIALS

#### ABSTRACT

Evidence Based Medicine (EBM) is generally defined as the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. In this section devoted to Evidence Based Paediatrics the conceptual, methodological and operational issues related to the practice of EBM in the field of Paediatrics have been covered. 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 randomised controlled trials (RCT's). A clinical scenario is presented and explored regarding the search and critical appraisal of scientific evidence about the efficacy of antibiotic therapy in children younger than two years with acute otitis media. Methods for searching the current best evidence are discussed and a systematic approach for the critical appraisal of RCT's is suggested, including: (1) assessment of validity and methodological quality of RCT's; (2) assessment of scientific and practical impact of study results and (3) assessment of their practical applicability.

**Key-words:** Randomised controlled trials, evidence based medicine, acute otitis media, antibiotic therapy, watchful waiting.

#### 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. Group E-BMW: 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, Richardson, W.S., Rosenberg, W., Haynes, R.B.: *Evidence Based Medicine*. London, U.K.: Churchill-Livingstone; 1997.
4. 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.
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. Damoiseaux RA, van Balen FA, Hoes AW, Verheij TJ, de Melker RA: Primary care based randomised, double blind trial of amoxicillin versus placebo for acute otitis media in children aged under 2 years. *Bmj* 2000, 320:350-354.
9. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T: The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001, 134:663-694.
10. 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.
11. Guyatt GH, Sackett DL, Cook DJ: Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *Jama* 1993, 270:2598-2601.
12. Guyatt GH, Sackett DL, Cook DJ: Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. *Jama* 1994, 271:59-63.
13. Dans AL, Dans LF, Guyatt GH, Richardson S: Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. Evidence-Based Medicine Working Group. *Jama* 1998, 279:545-549.
14. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA: Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-Based Medicine Working Group. *Jama* 1999, 282:771-778.
15. McAlister FA, Laupacis A, Wells GA, Sackett DL: Users' Guides to the Medical Literature: XIX. Applying clinical trial results B. Guidelines for determining whether a drug is exerting (more than) a class effect. *Jama* 1999, 282:1371-1377.
16. Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI: Stratified randomization for clinical trials. *J Clin Epidemiol* 1999, 52:19-26.
17. Schulz KF, Chalmers I, Grimes DA, Altman DG: Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *Jama* 1994, 272:125-128.
18. Schulz KF, Chalmers I, Hayes RJ, Altman DG: Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Jama* 1995, 273:408-412.

19. Pildal J, Hrobjartsson A, Jorgensen KJ, Hilden J, Altman DG, Gotzsche PC: Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol* 2007, 36:847-857.
20. Chalmers I: Assembling comparison groups to assess the effects of health care. *J R Soc Med* 1997, 90:379-386.
21. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP: Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998, 352:609-613.
22. Fergusson D, Aaron SD, Guyatt G, Hebert P: Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *Bmj* 2002, 325:652-654.
23. Soares I, Carneiro AV: Intention-to-treat analysis in clinical trials: principles and practical importance. *Rev Port Cardiol* 2002, 21:1191-1198.
24. Hollis S, Campbell F: What is meant by intention to treat analysis? Survey of published randomised controlled trials. *Bmj* 1999, 319:670-674.
25. Ruiz-Canela M, Martinez-Gonzalez MA, de Irala-Estevéz J: Intention to treat analysis is related to methodological quality. *Bmj* 2000, 320:1007-1008.
26. Schulz KF, Chalmers I, Altman DG: The landscape and lexicon of blinding in randomized trials. *Ann Intern Med* 2002, 136:254-259.
27. Schulz KF, Grimes DA: Blinding in randomised trials: hiding who got what. *Lancet* 2002, 359:696-700.
28. Boutron I, Estellat C, Guittet L, Dechartres A, Sackett DL, Hrobjartsson A, Ravaud P: Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. *PLoS Med* 2006, 3:e425.
29. Boutron I, Guittet L, Estellat C, Moher D, Hrobjartsson A, Ravaud P: Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med* 2007, 4:e61.
30. Haahr MT, Hrobjartsson A: Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. *Clin Trials* 2006, 3:360-365.
31. Fergusson D, Glass KC, Waring D, Shapiro S: Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. *Bmj* 2004, 328:432.
32. Fleming TR, DeMets DL: Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 1996, 125:605-613.
33. Montori VM, Jaeschke R, Schunemann HJ, Bhandari M, Brozek JL, Devereaux PJ, Guyatt GH: Users' guide to detecting misleading claims in clinical research reports. *Bmj* 2004, 329:1093-1096.
34. Bekelman JE, Li Y, Gross CP: Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *Jama* 2003, 289:454-465.
35. Lexchin J, Bero LA, Djulbegovic B, Clark O: Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *Bmj* 2003, 326:1167-1170.
36. Glasziou PP, Sanders SL: Investigating causes of heterogeneity in systematic reviews. *Stat Med* 2002, 21:1503-1511.
37. Bouchet C, Guillemin F, Briancon S: Nonspecific effects in longitudinal studies: impact on quality of life measures. *J Clin Epidemiol* 1996, 49:15-20.
38. Burneo JG, Montori VM, Faught E: Magnitude of the placebo effect in randomized trials of antiepileptic agents. *Epilepsy Behav* 2002, 3:532-534.
39. Shapiro AK: The placebo effect in the history of medical treatment: implications for psychiatry. *Am J Psychiatry* 1959, 116:298-304.
40. Shapiro AK: Factors Contributing to the Placebo Effect. Their Implications for Psychotherapy. *Am J Psychother* 1964, 18:SUPPL 1:73-88.
41. McCarney R, Warner J, Illife S, van Haselen R, Griffin M, Fisher P: The Hawthorne Effect: a randomised, controlled trial. *BMC Med Res Methodol* 2007, 7:30.
42. Brookes ST, Whitley E, Egger M, Smith GD, Mulheran PA, Peters TJ: Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol* 2004, 57:229-236.
43. Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G: Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol Assess* 2001, 5:1-56.
44. Rothwell PM: Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 2005, 365:176-186.
45. Guyatt G: *Users' guides to the medical literature : essentials of evidence-based clinical practice*. 2nd edn. New York McGraw-Hill Medical; 2008.
46. Bezakova N, Damoiseaux RA, Hoes AW, Schilder AG, Rovers MM: Recurrence up to 3.5 years after antibiotic treatment of acute otitis media in very young Dutch children: survey of trial participants. *Bmj* 2009, 338:b2525.
47. Glasziou PP, Del Mar CB, Sanders SL, Hayem M: Antibiotics for acute otitis media in children. *Cochrane Database Syst Rev* 2004:CD000219.
48. Koopman L, Hoes AW, Glasziou PP, Appelman CL, Burke P, McCormick DP, Damoiseaux RA, Le Saux N, Rovers MM: Antibiotic therapy to prevent the development of asymptomatic middle ear effusion in children with acute otitis media: a meta-analysis of individual patient data. *Arch Otolaryngol Head Neck Surg* 2008, 134:128-132.
49. Rovers MM, Glasziou P, Appelman CL, Burke P, McCormick DP, Damoiseaux RA, Gaboury I, Little P, Hoes AW: Antibiotics for acute otitis media: a meta-analysis with individual patient data. *Lancet* 2006, 368:1429-1435.
50. Vouloumanou EK, Karageorgopoulos DE, Kazantzi MS, Kapaskelis AM, Falagas ME: Antibiotics versus placebo or watchful waiting for acute otitis media: a meta-analysis of randomized controlled trials. *J Antimicrob Chemother* 2009, 64:16-24.