

AVALIAÇÃO CRÍTICA DE ARTIGOS RELATIVOS A ENSAIOS CLÍNICOS CONTROLADOS

I. Planeamento

C. SAMPAIO, J.J. FERREIRA

Instituto de Farmacologia e Terapêutica Geral. Faculdade de Medicina de Lisboa.
Serviço de Neurologia. Hospital de Santa Maria. Lisboa.

RESUMO

Neste trabalho descrevem-se, numa perspectiva pedagógica, vários aspectos relativos à fase de planeamento e execução dum ensaio clínico controlado destinado a comparar estratégias terapêuticas. Dão-se indicações sobre os aspectos críticos deste tipo de investigação clínica de forma a fornecer o leitor com os instrumentos que permitem determinar a validade dos resultados desse ensaio.

SUMMARY

Critical Assessment of Articles on Controlled Clinical Trials I. Planning

This is the first of two articles concerning the techniques of critical appraisal of controlled clinical trials. It presents an overview of the delicate points of the following issues: operational questions; definition of primary and secondary variables, classification of the trial as pragmatic or explanatory, selection of the population sample, study design and its appropriateness to the goals, sample size and overall statistical analysis. Particular attention is given to the value of randomisation and blinding for the prevention of bias.

INTRODUÇÃO

A discussão do conceito de *Medicina Baseada na Evidência*, já realizada nos artigos anteriores desta série, mostra que existe uma necessidade premente de ensinar e treinar os aspectos técnicos que permitem validar os resultados de um estudo de investigação clínica publicado. Nesta série abordou-se, ainda, a importância de existirem sínteses fidedignas da literatura publicada, de forma a rentabilizar o tempo de estudo e a tornar eficiente a procura de uma resposta para um problema clínico. A análise das fontes primárias de informação,

isto é, dos artigos originais, continua no entanto, a ser importante em várias circunstâncias: 1. Para os especialistas que participam activamente na investigação clínica da sua área e têm que se manter actualizados sem tempo de latência; 2. Quando os resultados da investigação clínica têm potencial para alterar rapidamente a prática clínica e, por isso, é urgente avaliar a sua validade; 3. Quando é necessária uma resposta para um problema para o qual não existem sínteses fidedignas; 4. Quando os resultados de uma determinada fonte primária são utilizados para induzir alterações na prática médica (notí-

cias veiculadas pelos meios de comunicação social; informações veiculadas pelos delegados de informação médica).

A análise das fontes primárias de informação depara sempre com um condicionante inultrapassável para quem é um simples leitor da literatura médica e não um investigador da qualidade dessa literatura. Um artigo científico é um relatório formal, estruturado e depurado da investigação que efectivamente foi feita e por vezes este relatório formal não descreve aspectos importantes da metodologia da investigação. Outras vezes, relata alguns aspectos utilizando conceitos que são inadequadamente aplicados. Deste facto resulta que a publicação respeitante a um determinado processo de investigação clínica é sempre uma representação da investigação que efectivamente foi feita e é sobre essa representação que a análise do leitor incide. Esta questão é particularmente pertinente em relação à fase de planeamento da investigação. A forma como a investigação foi planeada e executada é crucial para a validade dos seus resultados e por isso a leitura de uma publicação deve permitir exercer um juízo de valor sobre a qualidade metodológica destas fases da investigação clínica.

O objectivo da investigação clínica, quer ele seja a identificação de factores de risco para uma doença, a avaliação do valor diagnóstico de um novo método ou a avaliação da eficácia terapêutica de uma intervenção, determina as constricções metodológicas a que o planeamento e execução dessa investigação está sujeito. A avaliação da eficácia de uma intervenção terapêutica obriga à realização de ensaios clínicos controlados. Só em situações excepcionais estes poderão ser dispensados e nestes casos a valorização da eficácia será sempre pouco robusta.

Neste artigo pretendemos sintetizar algumas recomendações para avaliar a validade da fase de planeamento e execução da investigação em publicações relativas a ensaios clínicos.

AVALIAÇÃO CRÍTICA DA FASE DE PLANEAMENTO E EXECUÇÃO DE UM ENSAIO CLÍNICO PUBLICADO

Este tipo de avaliação crítica implica a valorização de diversos aspectos que são resumidos no quadro I e explicados abaixo.

1. A formulação da questão em avaliação.

No capítulo *Introdução* de qualquer artigo científico deve ficar claro qual é a pergunta ou perguntas a que o ensaio clínico pretende responder e quais são as razões pelas quais as referidas perguntas são consideradas perti-

Quadro I - Aspectos a valorizar na avaliação crítica

1. Formulação da questão em avaliação.
2. Definição das variáveis em análise: categorização em variáveis primárias e secundárias.
3. Caracterização do ensaio como pragmático ou explanatório.
4. Avaliação da selecção dos participantes (critérios de inclusão e exclusão).
5. Adequação da selecção dos doentes à caracterização do ensaio.
6. Descrição do desenho do estudo e adequação do desenho do estudo ao problema em avaliação.
7. Existência de aleatorização e adequação do método de aleatorização.
8. Comparabilidade dos grupos no início do ensaio.
9. Identidade dos tratamentos concomitantes nos grupos em comparação.
10. Existência de ocultação e características técnicas dessa ocultação.
11. Determinação da dimensão da amostra.
12. Programação da análise estatística incluindo a definição de subgrupos.

centes. A equação destas perguntas obedece ao mesmo tipo de recomendações enunciadas para a formulação de questões clínicas noutro artigo desta série. As perguntas devem ser operacionais e não genéricas. Por exemplo se se disser que se pretende saber se o medicamento X é útil no tratamento do cancro da mama estamos perante uma pergunta mal formulada porque não é operacional, ou seja, não é possível ao leitor saber qual é o critério de utilidade do autor nem qual o tipo ou qual o estadió de cancro da mama a que se está a referir. O critério de utilidade poder ser a redução da massa tumoral, aumento da sobrevida, aumento do período livre de recidivas, aumento da qualidade de vida, etc. A mesma pergunta correctamente formulada seria *pretende-se avaliar se o medicamento X aumenta a sobrevivência de mulheres pré-menopáusicas com cancro da mama metastizado, quando comparado com a intervenção Y.*

Num artigo em que não é possível encontrar uma questão correctamente operacionalizada, dificilmente se encontrarão os outros aspectos, considerados relevantes nesta análise, bem formulados. A operacionalização da questão é essencial para a definição das variáveis em estudo e da sua classificação em primárias e secundárias. Esta classificação é determinante da adequação dos instrumentos de medida (escalas de severidade de sintomas, exames radiológicos, escalas de qualidade de vida, contagem de eventos pré-definidos, entre outros). Por sua vez, a escolha das variáveis e dos instrumentos de medida é instrumental na determinação da dimensão da amostra.

2. Definição das variáveis em análise: categorização em variáveis primárias e secundárias.

A execução de um ensaio clínico, tal como qualquer outro projecto de investigação clínica, deve estar subordinado a uma questão correctamente formulada. Esta questão irá condicionar as variáveis, isto é, os parâmetros que tem interesse avaliar durante a execução do ensaio e cuja medida serve para responder à pergunta.

Existe sempre a tentação de rentabilizar o esforço da execução de um ensaio clínico multiplicando o número de parâmetros avaliados. Não caindo em exageros, ou seja, num número não maneável de parâmetros em análise, é possível conviver com esse desejo de rentabilizar o esforço investido. Para tanto, é necessário identificar claramente à partida quais são os parâmetros que são considerados variáveis primárias, isto é, aqueles que podem responder à questão inicialmente formulada e que devem ser muito poucos, um a dois. Por outro lado designamos as variáveis secundárias, isto é, aquelas que completam a resposta mas cuja análise é desprovida de sentido sem a resposta obtida através das variáveis primárias.

Por exemplo num ensaio destinado a avaliar uma associação de imunossupressores em comparação com uma outra associação de imunossupressores em doentes insuficientes renais transplantados, a variável primária pode ser a taxa de enxertos funcionantes ao fim de cinco anos e uma variável secundária a qualidade de vida dos doentes transplantados durante esses cinco anos.

A escolha das variáveis primárias irá determinar o cálculo da dimensão da amostra.

Numa perspectiva de avaliação crítica de um artigo publicado deve-se procurar sempre identificar as variáveis primárias e secundárias. Se esta identificação não for possível, ou for ambígua, isso pontuará em desfavor da validade dos resultados.

3. Caracterização do ensaio como pragmático ou explanatório.

Um ensaio clínico é pragmático ou explanatório consoante os seus objectivos e as suas variáveis primárias. Um ensaio clínico é designado pragmático quando pretende responder a uma pergunta cuja resposta têm imediata aplicabilidade clínica. Um ensaio pragmático é portanto um ensaio relevante para a avaliação do valor terapêutico de uma intervenção, independentemente do conhecimento do mecanismo pelo qual essa intervenção exerce o seu efeito. Um ensaio explanatório é aquele que tem por objectivo determinar o mecanismo de acção de uma intervenção. Os resultados de um ensaio expla-

natório podem estar muito distantes da aplicabilidade clínica. Aos ensaios pragmáticos corresponde o acrónimo anglo-saxónico POEM que significa *patient oriented end-points that matters* e aos ensaios explanatórios corresponde o acrónimo DOE que significa *disease oriented end-points*.

O facto de um ensaio ser pragmático ou explanatório não é em si mesmo positivo ou negativo, tudo depende da congruência com outros aspectos em avaliação. O carácter pragmático ou explanatório de um ensaio determina as características dos critérios de inclusão ou exclusão e a escolha das variáveis primárias. Se um ensaio se pretende pragmático os critérios de inclusão e exclusão devem ser pouco restritivos de forma a propiciar a aplicação dos resultados obtidos a uma população alargada; as variáveis primárias, num ensaio pragmático devem medir características com impacto directo no doente e não resultados laboratoriais intermediários.

Se um ensaio se diz pragmático e incide sobre uma amostra ultraseleccionada e/ou avalia parâmetros remotamente significativos para o doente, a avaliação crítica terá que ser negativa.

4. Avaliação da selecção dos participantes (critérios de inclusão e exclusão).

Em relação a esta questão o primeiro ponto a verificar é se estes critérios estão claramente descritos na secção dos métodos. Subsequentemente deve-se ajuizar se os critérios descritos são adequados para definir a população do estudo, tendo em atenção os objectivos deste. Ainda em relação a esta questão é também importante verificar se as características dos doentes efectivamente estudados estão de acordo com os critérios de selecção enunciados. Pode acontecer que, apesar destes critérios parecerem adequados, os doentes seleccionados representem apenas um subgrupo dos doentes permitidos ou tenham ocorrido violações ao protocolo.

5. Adequação da selecção dos doentes à caracterização do ensaio.

(ver alínea 3)

6. Descrição do desenho do estudo e adequação do desenho ao problema em avaliação.

Este é um ponto fulcral. Na avaliação de um ensaio clínico o desenho, isto é, a estrutura adoptada deve estar inequivocamente descrita. Deve ser claro se um estudo é uma comparação em paralelo, se é um estudo cruzado (*crossover*), se é um estudo factorial, etc. Os períodos de pré-selecção também devem estar perfeitamente referi-

dos. Por exemplo, se os doentes antes de serem distribuídos pelos grupos em confronto foram submetidos a um tratamento com placebo e se os doentes que tiveram uma resposta positiva ao placebo foram escolhidos (*run-in period*). A avaliação da existência de uma descrição é apenas a primeira fase. Posteriormente, é fundamental verificar se o desenho escolhido é apropriado à patologia em avaliação e aos objectivos. Por exemplo, os ensaios em cruzado só podem ser utilizados com patologias crónicas, relativamente estáveis ao longo do tempo e para tratamentos não curativos. Um estudo factorial só pode ser utilizado quando as diferentes intervenções em confronto não têm efeitos sinérgicos.

Também é importante avaliar o horizonte temporal do estudo, uma vez que a duração do período de acompanhamento é extraordinariamente relevante para a sua validade externa. Para uma doença crónica os resultados de uma terapêutica ao fim de três meses podem ser totalmente irrelevantes se não houver evidência de que são duráveis.

7. Existência e adequação do método de aleatorização.

A característica técnica mais importante de um ensaio clínico é a forma como são gerados os grupos em comparação. A aleatorização dos doentes pelos grupos em confronto, no caso dos ensaios em paralelo, ou da sequência dos tratamentos no caso dos ensaios em cruzado, é o único processo de assegurar que os investigadores por razões conscientes ou inconscientes, não enviesam a distribuição dos doentes ou dos tratamentos. Por outro lado, só a existência de aleatorização legítima a aplicação de testes estatísticos sobre os resultados obtidos.

Não basta que uma publicação relate que a distribuição dos doentes ou das intervenções foi conduzida por aleatorização; é necessário também que se descreva o processo e que este seja compatível com uma verdadeira ocultação da sequência de aleatorização. Por exemplo, a distribuição dos doentes de acordo com as suas datas de nascimento serem pares ou ímpares, ou por alternância não são verdadeiros métodos de aleatorização porque permitem ao investigador adivinhar a *sorte* do próximo doente antes deste ter sido seleccionado. Não se pense que este aspecto é um detalhe de um metodologista, pois demonstrou-se experimentalmente que quando um ensaio não é aleatorizado os resultados favorecem excessivamente a intervenção *melhor* e que quando a aleatorização é apenas uma pseudo-aleatorização os resultados também são inflacionados, embora não tão significativamente como na primeira situação. Para haver uma ordem

de grandeza podemos citar que, nos estudos referidos uma mesma intervenção, testada num ensaio verdadeiramente aleatorizado e num não aleatorizado, aparece beneficiada em cerca de 41%¹ neste último.

Resta acrescentar que o método mais fidedigno de aleatorização é designada por aleatorização central, em que o investigador contacta um centro à distância para obter a atribuição de grupo ou de tratamento para um doente específico já seleccionado.

Salienta-se que a ocultação da sequência de aleatorização é um aspecto distinto da ocultação dos tratamentos, este tem em vista o controlo de outro tipo de viés.

8. Comparabilidade dos grupos no início do ensaio.

Quando se deixa à sorte, através da aleatorização, a distribuição dos grupos em confronto corre-se o risco de obter grupos não comparáveis em relação a características fundamentais. Se essas características são conhecidas *à priori*, por exemplo a existência de hábitos tabágicos em doentes incluídos num ensaio cuja variável primária é a ocorrência de um acidente cardiovascular, existem métodos para obviar o problema como seja a aleatorização estratificada ou aleatorização por minimização. Se uma característica não for considerada fundamental para determinar a utilização destes métodos e se se verificar *à posteriori* que há um desequilíbrio na distribuição de algumas variáveis, por exemplo o número de doentes residentes em meio urbano *versus* em meio rural, só uma análise qualitativa permitirá ponderar a relevância desses achados.

9. Identidade dos tratamentos concomitantes nos grupos em comparação.

Deve-se verificar se os grupos em confronto foram tratados de forma equivalente. Por exemplo se se está a estudar uma técnica cirúrgica o treino das equipas de cirurgia deve ser idêntico. O grupo da intervenção inovadora não pode receber cuidados extra em relação ao grupo controlo. Por exemplo se, num ensaio de uma terapêutica cirúrgica *versus* uma terapêutica médica, o grupo operado receber posteriormente fisioterapia e o grupo tratado medicamente não, os resultados passam a referir-se à comparação cirurgia mais fisioterapia relativamente à terapêutica médica e não à comparação estabelecida nos objectivos.

Se os grupos receberem diferentes tratamentos concomitantes e não for possível sistematizá-los de forma operacional, os resultados obtidos no ensaio poderão não ser interpretáveis.

Uma forma de contornar parcialmente este problema,

pelo menos a introdução deliberada de diferentes tratamentos concomitantes é através da ocultação simples ou dupla das intervenções em causa.

10. Existência de ocultação e características técnicas dessa ocultação.

Ocultação significa que a distribuição dos tratamentos em confronto é desconhecida dos doentes, no caso da ocultação simples, e dos doentes e dos investigadores no caso da ocultação dupla. O objectivo desta estratégia é obviar à introdução de enviesamento dos resultados atribuível às expectativas, quer dos doentes quer dos investigadores. Este aspecto é significativo mas não tanto quanto a aleatorização de que falámos acima. Os resultados de um estudo conduzido sem ocultação, *versus* o mesmo estudo conduzido com ocultação inflacionam os resultados positivos em cerca de 17%², bastante menos que a ausência de aleatorização.

Está, portanto, demonstrado que a ocultação é menos importante do que a aleatorização. No entanto, se os autores reclamam que o estudo foi conduzido em ocultação, a forma como esta é conseguida deve estar claramente descrita. Não é equivalente utilizar um terceiro investigador para realizar observações em condições de ocultação ou realizar um estudo comparativo em que os dois tratamentos são indistinguíveis.

11. Determinação da dimensão da amostra.

Um ensaio clínico pode ser inconclusivo, isto é, não demonstrar diferença *significativa* entre os termos em comparação e o leitor incauto concluir daí que os tratamentos testados são equivalentes quando de facto não o são. Neste contexto, o ensaio clínico deve ser considerado como um teste diagnóstico e sobre ele deve-se ajuizar se tem capacidade para diagnosticar a diferença que se considera relevante. Esta capacidade de diagnóstico é a sensibilidade do ensaio clínico. Na prática, a razão mais frequente para não se encontrarem diferenças entre dois tratamentos em comparação é a insuficiente sensibilidade do ensaio, e não a equivalência das intervenções. Nesta perspectiva, quando se avalia criticamente um ensaio clínico é fundamental verificar se a amostra foi calculada, de forma a dotar o ensaio da sensibilidade adequada para determinar a diferença que se considera clinicamente importante.

No cálculo da amostra entram vários factores: a diferença procurada (dimensão do efeito), a sua incerteza (desvio padrão), o erro a (a probabilidade de se concluir por efeito da sorte que dois tratamentos são diferentes quando de facto não são) que os investigadores consi-

deram aceitável e o erro b (a probabilidade de se concluir, por efeito da sorte, que dois tratamentos são idênticos quando na realidade são diferentes) que os investigadores consideram aceitável.

12. Programação da análise estatística incluindo a definição de subgrupos.

A análise estatística deve ser definida antes do ensaio começar. É claro que o leitor não tem meios para avaliar se de facto isto aconteceu. No entanto, a coerência global da análise efectuada pode ser um indicador que de facto a planificação aconteceu. A descrição da análise estatística, na secção de material e métodos deve incluir não só os testes estatísticos utilizados mas também as *assumpções* realizadas, e o nível de erros admitidos bem como as correções tidas em consideração para estes erros, no caso de comparações múltiplas ou da realização de análises interinas.

Para um leitor com conhecimentos médios de estatística a utilização de um teste desconhecido deve levantar alguma suspeita, a menos que a escolha esteja adequadamente justificada.

Outro aspecto importante é a definição *a priori* das análises de subgrupos consideradas.

CONCLUSÃO

Os sociólogos da ciência têm descrito repetidamente o comportamento dos cientistas no seu dia-a-dia e as discrepâncias do mesmo face a forma de apresentação dos resultados da actividade científica^{3,4}. Sir Medawar, que não era sociólogo mas um cientista reputado disse essencialmente o mesmo ao chamar a atenção, num ensaio célebre⁵, para quão anti-natural era a fragmentação de um artigo científico em introdução, material e métodos, resultados e conclusões. Não pretendemos aqui juntar-nos ao coro dessas críticas, que chegam a pôr em causa a validade do conhecimento científico, até porque consideramos a estruturação dos documentos científicos essencial ao funcionamento da ciência e não partilhamos o ponto de vista relativista na filosofia do conhecimento. Tão só ambicionamos chamar a atenção do leitor desprevenido para as dificuldades de interpretação de um documento, o artigo científico, que se pressupõem isento de paixão e consequentemente rigoroso. Neste trabalho chamamos a atenção para diversos problemas que ameaçam a validade dos resultados de um ensaio clínico. Estes problemas podem surgir por ignorância técnica, ingenuidade, desleixo, ou por fraude. As suas consequências são idênticas no que se refere à fragilização dos resultados. Não devemos, no entanto, esquecer que os casos de fraude

científica têm crescido o que também tem consequências para toda a comunidade científica e para os consumidores da produção desta.

Como medida pedagógica e de garantia da qualidade um grande número de revista médicas de grande prestígio (JAMA, BMJ, Lancet) adoptaram um algoritmo a que os manuscritos relativos a ensaios clínicos devem obedecer, se quiserem ser considerados para publicação. Esse algoritmo designa-se por CONSORT⁶ e o seu conhecimento é recomendado tanto a investigadores como a leitores de ensaios clínicos.

REFERÊNCIAS

No caso deste artigo que tem essencialmente uma função pedagógica optamos por apresentar uma lista extremamente restrita de referências e adicionar uma bibliografia recomendada.

1. SHULTZ 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, 273:408-12
2. MUIR GRAY JA: Randomised controlled trials in Evidence-based healthcare. How to make health policy and management decisions. Churchill Livingstone 1997,78-86.
3. Scientific Knowledge. A Sociological analysis. Barnes B, Bloor D,

Henry J. The Athlone Press, 1996.

4. SCRIVEN M: The exact role of value judgments in Science. In Ethical issues in scientific research. An anthology. Erwin E, Gendin S, Kleiman L. Gerland Publishing Inc, 1994.29-50.
5. MEDAWAR P: Is the scientific paper a fraud? In the strange case of the spotted mice and other classic essays on science. Oxford University Press,1996:33-39.
6. BEGG C, CHO M, EASTWOOD S, HORTON R, et al: Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1994; 276(8):637-639.

BIBLIOGRAFIA RECOMENDADA

- FREIMAN JA, CHALMERS TC, SMITH H, KUEBLER RR: The importance of Beta, the type II error, and sample size in the design and interpretation of the randomised controlled trial. N Eng J Med 1977, 299:690-4.
- GREENHALGH T: How to read a paper. The basics of evidence based medicine. BMJ publishing group, 1997.
- OXMAN AD, SACKETT DL, GUYATT GH, AND EVIDENCE BASED MEDICINE WORKING GROUP: User's guide to the medical literature I. How to get started. JAMA 1993, 270(17): 2093-2095.
- GUYATT GH, SACKETT DL, COOK D AND EVIDENCE BASED MEDICINE WORKING GROUP: User's guide to the medical literature II. How to use an article about therapy or prevention. A) Are the results of the study valid?. JAMA 1993, 270(21): 2598-2601.
- GUYATT GH, SACKETT DL, COOK D AND EVIDENCE BASED MEDICINE WORKING GROUP: User's guide to the medical literature II. How to use an article about therapy or prevention. B) What where the results and will they help me in caring for my patients?. JAMA 1994, 271(1): 59-63.