

# Princípios de Análise Metodológica de Ensaio Terapêuticos. Parte I: Validade dos Resultados [34]

ISABEL SOARES \*, ANTÓNIO VAZ CARNEIRO \*\*

\* Serviço de Medicina I, Hospital Egas Moniz, Lisboa

\*\* Centro de Estudos de Medicina Baseada na Evidência da Faculdade de Medicina de Lisboa

Rev Port Cardiol 2002;21 (4): 457-469

## RESUMO

Escolher a estratégia óptima de tratamento para um doente é uma das questões mais importantes que os cardiologistas e os médicos em geral defrontam na sua prática clínica diária. Uma das melhores fontes de informação em terapêutica são os ensaios aleatorizados e controlados, que, apesar de algumas desvantagens, são considerados o *gold standard* da investigação médica sobre medidas terapêuticas e preventivas.

A qualidade metodológica dos ensaios terapêuticos é muito variada e – como já foi apresentado noutros artigos desta série – para que sejam úteis é fundamental a avaliação crítica da sua qualidade, em termos de validade dos seus resultados, da sua importância e da sua aplicabilidade ao doente individual.

No presente artigo apresentam-se as regras a que deve obedecer a avaliação crítica de estudos sobre intervenções terapêuticas, incidindo sobre a validade dos resultados. Em artigos posteriores, apresentaremos as regras de avaliação crítica em termos da importância e da aplicabilidade dos resultados.

## Palavras-Chave

Ensaio aleatorizado prospectivo e controlado;  
Validade Interna; Avaliação crítica;  
Cardiologia baseada na evidência

## ABSTRACT

### Basics of Methodologic Analysis of Therapeutic Trials. Part I: Validity of the Results

One of the most compelling questions a cardiologist (and, for that matter, any clinician) has to face involves choosing the optimal management strategy for their patients. One of the best sources of information about therapy is the randomized controlled trial, which, despite several significant shortcomings, is considered to be the gold standard in studies about therapeutics and prevention.

The methodological quality of therapeutic trials is highly variable and – as already discussed in previous articles – in order to be useful they have to be critically appraised in terms of the validity of their results, their importance, and their applicability to the individual patient.

In this article, we will introduce guidelines for the critical appraisal of evidence in therapy, discussing the issue of validity. In future articles we will present a critical appraisal of the importance and applicability of randomized clinical trial results.

## Key words

Randomised clinical trials;  
Internal validity;  
Critical approach; Evidence-based cardiology

## INTRODUÇÃO

Escolher a estratégia ótima de tratamento para um doente é uma das questões mais importantes que os cardiologistas e os médicos em geral defrontam na sua prática clínica diária. Por exemplo, quais são os benefícios de prescrever tratamento farmacológico ou alteração da dieta para baixar a tensão arterial, o nível de colesterol ou o peso de um doente? Que é que os doentes podem esperar, em termos de benefício sintomático ou aumento da longevidade, de tratamento da sua insuficiência cardíaca crónica, diabetes ou dependência do tabaco? Igualmente importante, que efeitos adversos a curto ou longo prazo são previsíveis em consequência da intervenção?

Nestas questões há dois aspectos relacionados<sup>(1)</sup>. Em primeiro lugar, a que riscos estão sujeitos os doentes que fumam ou têm excesso de peso, que são hipertensos, que têm valores elevados de colesterol ou glicose, ou uma função cardíaca anormal? Estes são os aspectos de dano/prejuízo. Em segundo lugar, no caso de ser instituída uma intervenção para modificar o comportamento dos doentes ou a fisiologia do seu organismo, quais são os benefícios que daí podem resultar, e em que medida estes benefícios ultrapassam quaisquer consequências nefastas? Estes são os aspectos de terapêutica.

Ao abordar estas questões estamos a defrontar problemas de causalidade. Num grupo particular de pessoas há uma relação causal entre uma exposição (fumo de tabaco, obesidade, tensão arterial elevada), ou uma intervenção (programa de cessação do tabagismo ou de perda de peso, fármaco anti-hipertensivo), e um determinado resultado antecipado (cancro do pulmão, enfarte do miocárdio, acidente vascular cerebral) ou não antecipado.

Para cada uma destas questões há uma resposta verdadeira subjacente. Se as nossas inferências a respeito desta verdade estão erradas, as consequências podem ser extremamente negativas. Aprender as regras da evidência que permitem diferenciar asserções válidas de falsas é a única forma de evitar actuações desastrosas, de que há inúmeros exemplos na história da Medicina<sup>(2)</sup>.

Ao explorar uma questão de tratamento, os investigadores têm muito mais controle do que quando abordam um problema de dano/prejuízo. Podem, por exemplo, determinar quem recebe a intervenção experimental e quem re-

## INTRODUCTION

One of the most compelling questions a cardiologist (and, for that matter, any clinician) has to face in daily clinical practice involves choosing the optimal management strategy for their patients. For example, what are the benefits of prescribing pharmacological treatment or change in diet to reduce a patient's blood pressure, cholesterol levels, or weight? What, in terms of improvement in symptoms or increased longevity, can patients expect from treatment for heart failure, diabetes or tobacco dependency? Equally important, what short- or long-term adverse effects can be predicted as a result of the intervention?

There are two related aspects to these questions<sup>(1)</sup>. Firstly, what are the risks to patients who smoke, are overweight, or who have hypertension, high cholesterol or glucose levels, or abnormal heart function? These are the aspects related to harm. Secondly, when an intervention is designed to modify patients' behavior or physiology, what are the benefits that may result, and to what extent do these benefits outweigh any harmful consequences? These are the therapeutic aspects.

In asking these questions we are dealing with problems of causality. In a given group of people there is a causal relationship between an exposure (tobacco smoke, obesity, high blood pressure) or an intervention (smoking cessation or weight loss programs, antihypertensive medication) and a particular outcome that is expected (lung cancer, myocardial infarction, stroke) or not expected.

For each of these questions there is an underlying true answer. If our conclusions with regard to this truth are wrong, the consequences can be extremely negative. The history of medicine is full of examples of disastrous interventions<sup>(2)</sup>, and these can only be avoided by applying the rules of evidence that enable valid assertions to be distinguished from false ones.

Investigators have much more control when exploring a question of treatment than when considering a question of harm. They may, for instance, decide who is given the experimental treatment and who is given the control treatment; they may allocate patients to the two groups in a process similar to that of tossing a coin (randomization), thus conducting a randomized controlled trial; they can also organize their study in such a way that neither patients nor clinicians know which patients are receiv-

cebe a intervenção de controlo; podem distribuir os doentes pelos dois grupos, de acordo com um processo semelhante ao lançamento de uma moeda ao ar (chamado aleatorização) conduzindo desta forma um ensaio aleatorizado e controlado; podem, além disso, organizar o seu estudo de forma a que nem os doentes nem os clínicos saibam quais são os doentes que recebem o tratamento experimental ou de controlo.

Pelo contrário, quando investigam uma questão de dano/prejuízo, não existe este tipo de controlo. Não é possível ditar às pessoas como é que elas devem comportar-se, nem esconder delas o ambiente em que vivem. Neste caso, os investigadores usam estudos observacionais: podem simplesmente seguir doentes que, em resultado de decisão pessoal ou circunstâncias casuais, foram expostos a um determinado estímulo lesivo. Seguem-nos para a frente no tempo (prospectivamente), para determinar se vêm a sofrer o resultado que os preocupa – o resultado alvo (estudos de *coorte*). Ou, em alternativa, os investigadores seleccionam doentes que já sofreram o resultado alvo, seleccionam ainda outro grupo que não o apresenta e fazem uma comparação para ver em que medida os dois grupos foram expostos ao agente lesivo que se supõe ser a causa (estudos caso-controle).

A selecção de um grupo de controlo eficaz é um dos problemas mais difíceis que os autores de um estudo observacional defrontam<sup>(3, 4)</sup>. Em poucos ou nenhuns estudos de *coorte* foi possível identificar dois grupos de pessoas iguais em termos de idade, sexo, *status* socio-económico, presença de doenças concomitantes, tendo como única diferença a exposição ao agente em estudo. Na prática, muito do «controlo» em estudos observacionais é feito na fase de análise, com ajustamentos estatísticos complexos para diferenças basais nas variáveis-chave<sup>(5)</sup>. Como resultado, as inferências de estudos que investigam dano/prejuízo são em geral muito mais fracas do que as que são possíveis de estudos que investigam terapêutica<sup>(1)</sup>. Os critérios que se aplicam a estudos que investigam uma questão terapêutica são diferentes dos que se aplicam a estudos que investigam uma exposição prejudicial.

Há, no entanto, excepções a estas regras gerais. Por um lado, a exposição lesiva pode por vezes ser uma intervenção médica (como um fármaco) e se o efeito prejudicial ocorrer rápida e frequentemente, pode ser usado o tipo de estudo habitualmente associado a interven-

ing the experimental treatment and which are controls.

By contrast, when a question of harm is being investigated, this type of control is not available. It is not possible to dictate how people should behave, nor conceal from them the environment in which they live. In this case, investigators use observational studies: they may simply follow patients who, through a personal decision or through force of circumstances, have been exposed to a particular harmful stimulus. They follow them over time (prospectively), to determine whether they will suffer the outcome under consideration, or target outcome (cohort studies). Alternatively, the investigators may select patients who have already suffered the target outcome, select another group who have not, and compare the two to find out to what extent the two groups have been exposed to the harmful agent that is presumed to be the cause (case-control studies).

Selecting an effective control group is one of the most difficult problems facing the authors of an observational study<sup>(3, 4)</sup>. In few, if any, cohort studies has it been possible to find two groups of people who are identical in terms of age, gender, socioeconomic status, and presence of concomitant disease, and who only differ in exposure to the agent under study. In practice, much of the “control” in observational studies is carried out at the analysis stage, with complex statistical adjustments for the baseline differences in key variables<sup>(5)</sup>. As a result, the conclusions of studies on harm are in general much weaker than those from studies on therapy<sup>(1)</sup>. The criteria applicable to studies investigating a question of treatment are therefore different from those for studies investigating exposure to harm.

There are, however, exceptions to these general rules. On the one hand, exposure to harm may sometimes result from a medical intervention, such as a drug, and if the harmful effect occurs rapidly and frequently, the type of study usually associated with therapeutic interventions can be used to assess the possibility that there is a causal relationship between the drug and the toxic effect. On the other hand, as regards therapeutic interventions, there are significant disadvantages associated with randomized trials, mainly linked to their costly and time-consuming nature<sup>(3)</sup>. This means that:

- many trials are never performed, or are performed with too few patients or over too short a period of time;

ções terapêuticas, para avaliar a possibilidade de existir uma relação causal entre o fármaco e o efeito tóxico. Por outro lado, no que diz respeito a intervenções terapêuticas, há importantes desvantagens associadas a ensaios aleatorizados, relacionadas em larga medida com o facto de serem caros e consumidores de tempo<sup>(3)</sup>.

Deste facto resulta que:

- muitos ensaios nunca são feitos, ou são feitos num número demasiado pequeno de doentes, ou por períodos demasiado curtos;
- muitos ensaios são financiados por grandes instituições de investigação ou laboratórios, que, em última instância, ditam a agenda de investigação;
- na avaliação dos efeitos, são muitas vezes usados marcadores substitutivos (*surrogate end points*) em vez de resultados clínicos «duros».

Há, além disso, muitas situações em que os ensaios aleatorizados e controlados são ou desnecessários, ou impraticáveis, ou inapropriados. Não é possível fazer ensaios aleatorizados para investigar condições raras, a qualidade do cuidado dentro de um hospital ou entre diferentes hospitais, ou intervenções na comunidade<sup>(6)</sup>.

Quando um médico, que está perante a necessidade de tomar uma decisão terapêutica, verifica que não há ensaios aleatorizados e controlados, deve apoiar-se em estudos de coorte e caso-controle – neste caso, a evidência mais forte disponível. Ao fazer isto, deve aplicar os critérios apropriados para a avaliação destes estudos, critérios que são em regra associados com a investigação de exposições potencialmente prejudiciais. Deve também ter sempre presente que a força de quaisquer inferências acerca da relação causal entre a intervenção e o resultado é muito mais fraca do que seria se a evidência viesse de um ensaio aleatorizado<sup>(7)</sup>.

Vamos de seguida expor as regras a que deve obedecer a avaliação crítica de estudos sobre intervenções terapêuticas: neste texto, o termo «terapêutica» será usado num sentido amplo. Será importante recordar que os mesmos critérios podem ser aplicados na avaliação de estudos sobre intervenções terapêuticas (dirigidas para reduzir sintomas e curar doença) ou sobre intervenções preventivas (dirigidas para reduzir o risco de doença ou complicações de doença).

- many trials are funded by large research institutions or laboratories, which ultimately dictate the research agenda;
- in assessment of the effects, substitute markers or “surrogate endpoints” are often used instead of ‘hard’ clinical results.

There are, moreover, many situations in which randomized controlled trials are unnecessary, or impractical, or inappropriate. Randomized trials cannot be performed to investigate rare conditions, to evaluate the quality of care within a particular hospital or between hospitals, or to assess community-based care<sup>(6)</sup>.

When a physician, faced with the need to make a therapeutic decision, finds that there are no relevant randomized controlled trials, he or she should base their decision on cohort and case-control studies – in this case, the best available evidence. For this, the appropriate criteria should be applied in assessing these studies, such criteria generally being associated with the investigation of potentially harmful exposure. It should also always be borne in mind that any inference regarding a causal relationship between the intervention and the outcome is much weaker than if the evidence had come from a randomized trial<sup>(7)</sup>.

Below we set out the guidelines to be followed in the critical appraisal of studies on therapeutic interventions. The term ‘therapy’ will be used in a wide sense; it is important to remember that the same criteria can be applied for appraisal of studies on therapeutic interventions (designed to reduce symptoms and cure disease) and on preventative intervention (intended to reduce the risk of disease or associated complications).

## APPRAISAL OF STUDIES ON THERAPY

When consulting a medical article to try and resolve a particular clinical problem, including a question of treatment, the best approach involves three successive steps, as described in *Table I*:

1. Are the results of the trial valid? This relates to the credibility of the results and is intended to confirm whether the assessment of the treatment effect accurately represents the nature and magnitude of the real underlying effect. The question may be posed in another way: do

## AVALIAÇÃO DE ESTUDOS SOBRE TERAPÊUTICA

Ao utilizar um artigo médico para tentar resolver qualquer problema clínico, inclusive terapêutico, deve-se seguir uma abordagem que envolve três degraus sucessivos<sup>(8)</sup> e que estão descritos no *Quadro I*:

1. Os resultados do estudo são válidos? Esta questão tem a ver com a credibilidade dos resultados e tem por objectivo averiguar se a avaliação do efeito do tratamento representa de um modo fidedigno a direcção e magnitude do verdadeiro efeito subjacente. Pode ser colocada de outro modo: estes resultados representam uma avaliação imparcial do efeito do tratamento, ou foram influenciados de uma forma sistemática de modo a induzir uma falsa conclusão? Se os resultados são válidos, vale então a pena prosseguir na sua análise.
2. Quais são os resultados? Esta questão considera a amplitude e precisão do efeito do tratamento. A melhor avaliação da amplitude do efeito do tratamento são os próprios resultados do estudo, sendo que a precisão da avaliação é superior em estudos com maior número de doentes.
3. Estes resultados são aplicáveis ao meu doente? Esta questão tem duas partes: em primeiro lugar, é necessário saber se os resultados são generalizáveis ao doente concreto, ou se o doente é demasiado diferente dos que participaram no ensaio para haver dúvidas quanto ao interesse do tratamento específico; em segundo lugar, se os resultados são aplicáveis, qual é o impacto líquido do tratamento. O impacto depende tanto dos benefícios como dos riscos (efeitos colaterais e efeitos tóxicos) do tratamento, mas também das consequências de não administrar o tratamento em questão. Assim, pode decidir-se não administrar uma terapêutica que é efectiva nos casos em que o prognóstico do doente já é bom sem tratamento, especialmente quando há o risco de haver importantes efeitos laterais e toxicidade.

Neste artigo, iremos apresentar apenas as regras de avaliação crítica referentes à primeira questão:

these results represent an impartial assessment of the treatment effect, or were they systematically influenced, thus leading to a false conclusion? If the results are valid, it is worth continuing with the analysis.

2. What are the results? This question involves the extent and precision of the treatment effect. The best assessment of the extent of the treatment effect is in the actual results of the trial, and the greater the number of patients, the greater the accuracy of the assessment.
3. Are these results applicable to my patient? This question is in two parts. Firstly, it is necessary to ascertain whether the results can be generalized to a particular patient, or whether the patient is so different from those who participated in the trial that there are doubts about the relevance of the specific treatment. Secondly, if the results are applicable, what is the net impact of the treatment? The impact depends on the benefits as well as on the risks (side and toxic effects) of the treatment, but also on the consequences of not giving the treatment in question. It may thus be decided not to administer an effective therapy in cases where the patient's prognosis is good with no treatment, especially when there is a risk of significant side effects or toxicity.

*Table I*

### Critical appraisal of studies

---

Are the results of the trial valid?  
What are the results?  
Are these results applicable to my patient?

---

In this article, we shall present the guidelines for critical appraisal with reference to the first question only.

### ARE THE RESULTS VALID?

The validity of the results of a trial depends on how it was designed and conducted, in other words on the methodology that was used. The design is one of the most important aspects of a trial, since it plays a crucial role in determining causality.

The analysis of the methodology of a therapeutic trial is set out in *Table II* and involves two sets of four questions<sup>(1, 9, 10)</sup>. The first set is

#### Quadro I

##### Avaliação crítica de estudos

Os resultados do estudo são válidos?  
Quais são os resultados?  
Estes resultados são aplicáveis ao meu doente?

### OS RESULTADOS SÃO VÁLIDOS?

A validade dos resultados de um estudo resulta do modo como ele foi desenhado e conduzido, isto é, da metodologia que foi usada. O *design* constitui uma das características mais importantes de um estudo, já que tem uma importância crucial na determinação da causalidade.

A análise metodológica de um estudo sobre terapêutica está descrita no *Quadro II* e utiliza dois grupos de quatro questões<sup>(1, 9, 10)</sup>: o 1.º grupo tem por objectivo averiguar se no início do estudo os doentes expostos à intervenção experimental tinham um prognóstico semelhante aos doentes de controlo; o 2.º grupo tem por objectivo confirmar que os dois grupos permaneceram semelhantes, no que diz respeito a factores de prognóstico, ao longo de todo o estudo.

#### Quadro II

##### Avaliação crítica da validade dos resultados de um ensaio terapêutico

1. Os grupos experimental e controlo começaram o estudo com um prognóstico semelhante?
  - 1.1. Os doentes foram aleatorizados?
  - 1.2. A aleatorização foi ocultada?
  - 1.3. Os doentes foram analisados nos grupos para os quais foram de início aleatorizados?
  - 1.4. Os doentes dos grupos experimental e controlo eram semelhantes no que diz respeito a factores de prognóstico?
2. Os grupos experimental e controlo mantêm um prognóstico semelhante depois do estudo ter começado?
  - 2.1. Os doentes sabiam a que grupo pertenciam?
  - 2.2. Os clínicos sabiam quais eram os grupos?
  - 2.3. Os avaliadores de resultados sabiam quais eram os grupos?
  - 2.4. O *follow-up* foi completo?

### 1. OS GRUPOS EXPERIMENTAL E CONTROLO COMEÇARAM O ESTUDO COM UM PROGNÓSTICO SEMELHANTE?

#### 1.1. Os doentes foram aleatorizados?

De todas, esta é a questão mais importante a colocar. A distribuição dos doentes pelos

intendidos para ascertainment, no início do estudo, os doentes expostos à intervenção experimental tinham um prognóstico semelhante aos doentes de controlo; o propósito do segundo conjunto de questões é confirmar que os dois grupos permaneceram semelhantes em termos de factores prognósticos durante o estudo.

#### Table II

##### Critical appraisal of the validity of results of a therapeutic trial

1. Did the experimental and control groups begin the study with a similar prognosis?
  - 1.1. Were the patients randomized?
  - 1.2. Was the randomization blinded?
  - 1.3. Were the patients analyzed in the groups to which they were initially randomized?
  - 1.4. Were the patients in the experimental and control groups similar with regard to prognostic factors?
2. Did the experimental and control groups maintain a similar prognosis after the trial began?
  - 2.1. Did the patients know which groups they belonged to?
  - 2.2. Did the clinicians know which group was which?
  - 2.3. Did those who analyzed the results know which group was which?
  - 2.4. Was follow-up complete?

### 1. DID THE EXPERIMENTAL AND CONTROL GROUPS BEGIN THE STUDY WITH A SIMILAR PROGNOSIS?

#### 1.1. Were the patients randomized?

This is the most important question of all. Were the patients distributed between the treatment groups in a random fashion? Did the investigators use a method analogous to tossing a coin to allocate the patients to their treatment groups?<sup>(11)</sup>

The importance of randomization lies in the fact that it is the best way to create groups of patients who, at the beginning of the trial, are identical in terms of the risk of suffering the event that the therapy is intended to prevent<sup>(12)</sup>. Randomized distribution ensures that the groups are evenly balanced with regard to prognostic factors which, if they were unequally distributed between the treatment groups, could exaggerate, nullify or reverse the treatment effects.

Clinical results have many causes, of which treatment is only one; the severity of the underlying disease, the presence of comorbid conditions, and a range of other prognostic factors, known and unknown, often destroy any effect

grupos de tratamento foi aleatorizada? Os investigadores usaram algum método análogo ao «lançar de uma moeda ao ar» para distribuir os doentes pelos grupos de tratamento?<sup>(11)</sup>

A importância da aleatorização resulta de que esta é a maneira mais adequada de criar grupos de doentes, no início do ensaio, que são idênticos no risco de sofrer o evento que queremos prevenir<sup>(12)</sup>. A distribuição aleatorizada equilibra os grupos em relação a factores de prognóstico que, se fossem desigualmente distribuídos entre os grupos de tratamento, podiam exagerar, anular ou contrariar os efeitos da terapêutica.

Os resultados clínicos resultam de muitas causas, e o tratamento é apenas uma delas: a gravidade da doença subjacente, a presença de condições comórbidas e um conjunto de outros factores de prognóstico (conhecidos e desconhecidos) muitas vezes varrem qualquer efeito da terapêutica. O nome dado a este tipo de características dos doentes (que são estranhas à questão posta) é de confundidor: os factores confundidores podem, ao distribuírem-se desigualmente entre os grupos experimental e de controlo, influenciar os eventos clínicos que estamos a tentar impedir com o tratamento<sup>(9)</sup>.

Se estes factores exagerassem os efeitos aparentes de um tratamento de outro modo ineficaz, os efeitos deste desequilíbrio poderiam levar a uma conclusão (falsa-positiva) de que o tratamento era útil quando de facto não era. Pelo contrário, se eles escondessem ou contrariassem os efeitos de um tratamento realmente eficaz, o seu desequilíbrio podia levar a conclusões (falsas-negativas) de que um tratamento útil era inútil ou mesmo prejudicial.

Embora haja outros modos de evitar confundimento - exclusão, amostragem estratificada, *matching*<sup>(11)</sup> - todos exigem que se saiba antecipadamente qual é o confundidor. A aleatorização equilibra os grupos de tratamento em relação a estes factores, mesmo quando não sabemos o suficiente a respeito da doença alvo para saber quais são eles todos.

Pode no entanto acontecer que, por mero acaso, a aleatorização não tenha êxito no seu objectivo de conseguir grupos com prognóstico semelhante (mesmo não havendo erros na sua execução): a probabilidade disto acontecer é maior em ensaios com pequeno número de doentes.

## 1.2. A aleatorização foi ocultada?

É também importante averiguar se a aleatorização foi ocultada dos médicos que trouxe-

of the treatment. Such characteristics of the patient, which are unconnected to the question being posed, are known as confounding factors, and, if distributed unequally between the experimental and control groups, may influence the clinical events that we are attempting to prevent with the treatment<sup>(9)</sup>.

If these factors exaggerate the apparent effects of an otherwise ineffective treatment, the effect of this imbalance could lead to the conclusion that the treatment was useful when in fact it was not (false positive). If, on the other hand, the confounding factors hide or negate the effects of a genuinely effective treatment, this imbalance could result in the conclusion that a useful treatment was useless or even harmful (false negative).

Although there are other ways to avoid confounding - exclusion, stratified sampling, matching<sup>(11)</sup> - they all require that the confounding factor should be known in advance. Randomization balances the treatment groups with regard to these factors, even when not enough is known about the target disease to be sure what they all are.

It can, of course, happen that, by pure chance, randomization fails in its purpose of achieving groups with a similar prognosis, even when done correctly. The probability of this occurring is greater in trials with a small number of patients.

## 1.2. Was the randomization blinded?

It is also important to ask whether the randomization was blinded - concealed from the physicians who proposed the patients for the trial<sup>(8, 13)</sup>. If the clinicians who decided that a particular patient was eligible were aware of which arm of the study (control or experimental) the patient would be allocated to - if, in other words, the randomization was not blinded - they could systematically distribute more or less seriously ill patients to either of the two groups. Doing so will defeat the purpose of the randomization and the trial's results will be biased<sup>(14)</sup>. If the randomization process is concealed, the physicians will not know which treatment the next patient will receive and will thus be unable to skew the distribution between the groups being compared, either consciously or unconsciously. Blinding the randomization is ensured by entrusting this task to an investigator who is not involved in the enrollment of participants in the trial, but often it is not made clear in articles whether or not the random-

ram os doentes para o ensaio<sup>(8, 13)</sup>: se os clínicos que tomam a decisão acerca da elegibilidade de um doente têm conhecimento do braço do estudo (controlo ou experimental) para onde o doente vai ser distribuído – por outras palavras, se a aleatorização não é ocultada – eles podem sistematicamente distribuir os doentes mais ou menos graves por qualquer dos dois grupos. Esta actuação destrói o objectivo da aleatorização e o estudo vem a fornecer resultados enviesados<sup>(14)</sup>. Se a aleatorização for encoberta os médicos não saberão que tratamento o próximo doente vai receber e assim serão incapazes de, consciente ou inconscientemente, distorcer o equilíbrio entre os grupos em comparação. A ocultação da aleatorização consegue-se fazendo com que este processo seja da responsabilidade de um investigador que não esteja envolvido no recrutamento dos participantes no ensaio, mas muitas vezes não é clarificado nos artigos se a lista de aleatorização foi ocultada ou não. Se a aleatorização foi feita pelo telefone, ou por algum outro método a distância do sítio onde os doentes estavam a entrar no ensaio (aleatorização remota), pode-se estar certo que foi.

Dada a importância da aleatorização, uma conclusão óbvia é que se o estudo não foi aleatorizado, não vale a pena ser lido. Só se não houver ensaios aleatorizados é que se deve voltar eventualmente a ele.

### **1.3. Os doentes foram analisados nos grupos para os quais foram de início aleatorizados?**

Em ensaios aleatorizados, os doentes nem sempre seguem as instruções de prescrição medicamentosa, quer por esquecimento, quer por recusa formal.

As razões porque os participantes não tomam a sua medicação estão muitas vezes relacionadas com o prognóstico<sup>(15)</sup>. Tem-se verificado repetidamente que os doentes que tomam e não tomam a medicação do estudo têm resultados muito diferentes, mesmo quando essa medicação é um placebo. Em vários ensaios randomizados, os doentes que não aderiram aos seus regimes de tratamento tiveram pior evolução do que os que tomaram a sua medicação de acordo com as instruções<sup>(16)</sup>. Excluir da análise os doentes que não aderem à terapêutica deixa assim para trás os que podem ser destinados a ter um melhor resultado (com as óbvias consequências nos resultados do estudo).

ization was blinded. If the randomization was done by telephone, or by some other method at a distance from the place where the patients were enrolled in the trial (remote randomization), one can be sure that it was.

Given the importance of randomization, an obvious conclusion is that if a trial was not randomized, it is not worth reading, except possibly if no relevant randomized trials exist.

### **1.3. Were the patients analyzed in the groups to which they were initially randomized?**

In randomized trials, patients do not always follow the instructions for their prescriptions, whether through forgetfulness or through outright refusal.

The reasons for participants not taking their medication are frequently related to their prognosis<sup>(15)</sup>. It has been found again and again that patients who take their medication have very different outcomes from those who do not, even when the medication is a placebo. In various randomized trials, patients who did not comply with their treatment regimes have had worse outcomes than those who took their medication in accordance with instructions<sup>(16)</sup>. Excluding from the analysis those patients who do not comply with the treatment thus leaves out those who might have had a better outcome, with obvious consequences for the trial results.

For these reasons, all patients (even those who did not take their medication or who, accidentally or deliberately, received the wrong treatment) should be analyzed in the groups to which they were initially allocated. This principle results in an “intention-to-treat” analysis, in which the analysis of outcomes is based on the treatment arm to which the patients belonged following randomization, and not on the treatment that they actually received. This strategy preserves the value of the randomization: prognostic factors, both known and unknown, will on average be distributed equally between the two groups, and the effect seen will be only that due to the treatment.

### **1.4. Were the patients in the experimental and control groups similar with regard to prognostic factors?**

The purpose of randomization is to form groups with the same prognosis with regard to

Por estas razões, todos os doentes (mesmo aqueles que não tomaram a sua medicação ou que, acidentalmente ou intencionalmente, receberam o tratamento errado), devem ser analisados nos grupos a que inicialmente foram atribuídos. Este princípio resulta numa análise de tipo intenção-de-tratar, que é uma análise de resultados baseada no braço de tratamento a que os doentes pertenciam em consequência da aleatorização, e não no tratamento que eles realmente receberam. Esta estratégia preserva o valor da aleatorização: os factores de prognóstico conhecidos e os não conhecidos serão em média igualmente distribuídos pelos dois grupos, e o efeito que vemos será exactamente o que é devido ao tratamento.

#### **1.4. Os doentes dos grupos experimental e controle eram semelhantes no que diz respeito a factores de prognóstico?**

O objectivo da aleatorização é criar grupos com prognóstico idêntico relativamente ao resultado alvo. Para que isto aconteça, os grupos de tratamento e de controle devem ser semelhantes em todos os factores que determinam os resultados clínicos de interesse, excepto num deles: a terapêutica que recebem. A garantia de que é isto que acontece é dada pela apresentação das características de prognóstico basais potencialmente relevantes das duas populações. O exame destas características permite verificar como funcionou a aleatorização. Se os factores de prognóstico conhecidos são bem equilibrados, a probabilidade dos factores de prognóstico desconhecidos serem também equilibrados é maior. Por vezes a aleatorização falha em conseguir este objectivo. Quando os grupos são pequenos o acaso pode causar desequilíbrios: quando o tamanho da amostra aumenta, é cada vez menos provável que isso aconteça<sup>(17)</sup>.

O problema não está em saber se há diferenças estatisticamente significativas nos factores de prognóstico conhecidos entre os grupos de tratamento (num ensaio aleatorizado sabe-se antecipadamente que quaisquer diferenças que ocorram, aconteceram por acaso), mas antes qual é a amplitude destas diferenças. Se são grandes, a validade do estudo pode estar comprometida. Quanto mais forte for a relação entre os factores de prognóstico e o resultado, e mais pequeno o ensaio, mais as diferenças entre os grupos enfraquecem a força de qualquer

the target outcome. For this to happen, the treatment and control groups must be similar with respect to all the factors that determine the clinical results under study, except one: the treatment that they receive. What ensures that this will happen is that the potentially relevant baseline prognostic characteristics of the two populations should be presented. Examination of these characteristics shows how well the randomization worked. If the known prognostic factors are well balanced, there is a greater likelihood that unknown prognostic factors will also be balanced. Sometimes randomization fails to achieve this objective. When the groups are small, chance can cause imbalances; the larger the sample size, the less likely this is to occur<sup>(17)</sup>.

The problem is not so much to know whether there are statistically significant differences in the known prognostic factors between the treatment groups (in a randomized trial it is known in advance that any differences are due to chance), but rather to know the magnitude of these differences. If they are large, the trial's validity may be compromised. The stronger the relation between prognostic factors and the result, and the smaller the trial, the more the differences between the groups weaken the validity of any inference about the efficacy of the treatment. However, all is not lost if the treatment groups are not the same at the baseline level. There are statistical techniques that enable trial results to be adjusted for baseline differences.

Documentation should be sought in the study on the relevant baseline characteristics of the two groups, generally presented in the form of a table, and if there are substantial differences, it should be ascertained whether the investigators carried out an analysis to adjust for these differences. When non-adjusted as well as adjusted analyses arrive at the same conclusion, we can have greater confidence in the validity of the trial results<sup>(18)</sup>.

## **2. DID THE EXPERIMENTAL AND CONTROL GROUPS MAINTAIN A SIMILAR PROGNOSIS AFTER THE TRIAL BEGAN?**

### **2.1. Did the patients know which groups they belonged to?**

Patients progress better if they believe that they are receiving treatment, even when such treatment is an inactive substance; the placebo effect is a well-known source of bias in trials<sup>(19)</sup>.

inferência acerca da eficácia da intervenção. No entanto, nem tudo está perdido se os grupos de tratamento não são semelhantes a nível basal. Há técnicas estatísticas que permitem o ajustamento dos resultados do estudo para diferenças basais.

Deve-se procurar no estudo documentação sobre as características basais relevantes dos dois grupos, em regra apresentada na forma de um quadro, e se existem diferenças substanciais deve-se verificar se os investigadores conduziram uma análise que ajustasse para estas diferenças. Quando tanto análises não ajustadas como ajustadas chegam à mesma conclusão, podemos ganhar confiança na validade dos resultados do estudo<sup>(18)</sup>.

## **2. OS GRUPOS EXPERIMENTAL E CONTROLE MANTEM UM PROGNÓSTICO SEMELHANTE DEPOIS DO ESTUDO TER COMEÇADO?**

### **2.1. Os doentes sabiam a que grupo pertenciam?**

Os doentes evoluem melhor se pensam que estão a receber um tratamento, mesmo quando esse tratamento é uma substância inactiva: o efeito placebo é um potencial viés largamente reconhecido em ensaios<sup>(19)</sup>.

Um placebo é um tratamento inactivo que é dado de forma a que o doente não sabe se está a receber ou não o tratamento activo. Num ensaio, o grupo de controlo – em significativo número de vezes – recebe um tratamento placebo, pelo que este efeito é igual em ambos os grupos. Mesmo na ausência de efeito placebo, os doentes podem responder a questões ou executar testes funcionais de modo diferente, conforme acreditam ou não que estão a tomar a medicação activa<sup>(20)</sup>. Por estas razões, os doentes em ambos os grupos de um ensaio devem desconhecer se estão ou não a receber a terapêutica activa. Chama-se a isto *blinding* ou ocultação.

Para conseguir este objectivo, no ensaio de um fármaco, os doentes do grupo controlo devem receber um comprimido ou cápsula inerte que seja idêntico em cor, sabor e consistência, à medicação activa, administrada aos doentes do grupo tratamento. Estes placebos asseguram que os doentes do grupo controlo beneficiam de efeito placebo na mesma medida que os doentes activamente tratados<sup>(21)</sup>.

A placebo is an inactive treatment which is given so that the patient does not know whether he or she is receiving the active treatment. In a trial, the control group often receives a placebo, so that the effect is the same in both groups. Even when a placebo effect is not present, patients may respond to questions or perform functional tests differently depending on whether they believe that they are taking an active medication<sup>(20)</sup>. For these reasons, patients in both groups in a trial should not know whether they are receiving the active treatment or not. This is called blinding.

To achieve this objective in a drug trial, the patients in the control group should receive an inert tablet or capsule that is identical in color, taste and consistency to the active medication that is administered to patients in the treatment group. These placebos will ensure that the patients in the control group benefit to the same effect from the placebo effect as the patients in the actively treated group<sup>(21)</sup>.

### **2.2. Did the clinicians know which group was which?**

If the randomization is effective, the experimental (treatment) group and the control group begin the trial with very similar prognoses. However, randomization does not guarantee that the two groups will remain balanced in terms of prognosis. If there are differences in the treatment given to each of the groups (apart from the intervention under study), the results may be biased<sup>(9)</sup>.

The results of a trial are more trustworthy when the investigators can show that all additional treatments that could influence the results (cointerventions) were administered identically to both groups. Lack of such evidence is not a serious problem so long as the investigators do not know which patients are receiving active treatment and which are part of the control group (blinding). Proper blinding eliminates the possibility of the conscious or unconscious differential administration of effective cointerventions to the treatment and control groups<sup>(22)</sup>.

### **2.3. Did those who analyzed the results know which group was which?**

If the result involves clinical judgement and the trial investigators responsible for the assessment know that the patient was receiving the experimental treatment, their interpretation

## 2.2. Os clínicos sabiam quais eram os grupos?

Se a aleatorização resulta, os grupos experimental (de tratamento) e de controlo entram no estudo com um prognóstico muito semelhante. No entanto, a aleatorização não fornece garantias de que os dois grupos permaneçam equilibrados em relação ao prognóstico. Se há diferenças nas intervenções que são administradas a cada um dos grupos (para além da intervenção que está a ser estudada), os resultados podem vir enviesados<sup>(9)</sup>.

Os resultados de um estudo merecem mais confiança quando os investigadores documentam que todos os tratamentos adicionais que podem influenciar os resultados (co-intervenções) são administradas de modo idêntico a ambos os grupos. A ausência de tal documentação não é um problema grave se os investigadores desconhecem que doentes estão a receber tratamento activo e que doentes são parte do grupo controle (*blinding* ou ocultação). A ocultação efectiva elimina a possibilidade de administração diferencial, consciente ou inconsciente, de co-intervenções eficazes aos grupos tratamento e controlo<sup>(22)</sup>.

## 2.3. Os avaliadores de resultados sabiam quais eram os grupos?

Se o resultado implica julgamento clínico e os colaboradores do ensaio responsáveis pela avaliação sabem que o doente esteve a fazer o tratamento experimental, a sua interpretação dos resultados pode ser exposta a viés. Um avaliador *unblinded* (que sabe quais os grupos e tratamentos) que está a medir ou a registar resultados – como testes fisiológicos, *status* clínico ou qualidade de vida - pode fornecer diferentes interpretações de achados marginais, ou pode oferecer encorajamento diferencial durante a realização de testes, cada um dos quais pode distorcer os resultados<sup>(23)</sup>.

É extremamente difícil ser-se totalmente imparcial na avaliação clínica de um resultado, quando se está a par da intervenção que o doente sofreu. Isto também se estende à avaliação de achados clínicos registados por outros. Num ensaio, os avaliadores de resultados quase sempre podem ser mantidos «cegos», mesmo se, como é o caso de terapêuticas cirúrgicas, o doente e os clínicos não podem.

Na prática, quanto mais julgamento clínico é necessário para saber se um doente sofreu um resultado alvo, mais importante a ocultação se torna.

of the results may be subject to bias. An “unblinded” assessor (one who knows which group had which treatment) who is measuring or recording results, such as physiological tests, clinical status or quality of life, may put different interpretations on marginal findings, or may offer different levels of encouragement during the tests, either of which may distort the results<sup>(23)</sup>.

It is extremely difficult to be totally impartial in the clinical assessment of a result when one knows what treatment the patient has received. This is also true of the assessment of clinical findings recorded by others. In a trial, the assessors of outcomes can almost always be kept “blind”, even if, as in the case of surgical interventions, the patient and the physician cannot.

In practice, the more clinical judgement that is required to know whether a patient has had a target outcome, the more important blinding becomes.

## 2.4. Was follow-up complete?

Ideally, all the patients enrolled in a trial should be referenced at its conclusion.

If this does not happen, or if substantial numbers of patients are lost to follow-up, the validity of the trial may be called into question, the degree to which the trial results are compromised being proportional to the number of patients lost. This is because the lost patients frequently have a different prognosis from those who remain in the trial. Patients may disappear because they have had adverse outcomes, or because they are progressing well and fail to keep their appointments. The situation is the same as one that calls for an intention-to-treat analysis: patients who suspend their medication may have a smaller or (more often) a greater likelihood of suffering the adverse target event<sup>(24)</sup>.

When reading an article, the clinician can check for him- or herself whether the loss to follow-up was excessive by assuming, in a positive trial, that all the patients lost from the treatment group had an adverse outcome and that all those lost from the control group had a good outcome (the worst-case scenario) and by calculating the results again on the basis of these assumptions<sup>(1, 8)</sup>: if the trial conclusions remain unchanged, then loss to follow-up was not excessive. If they do change, the validity of the inferences is reduced, the degree of weakening depending on the likelihood of the patients

## 2.4. O follow-up foi completo?

Idealmente, todos os doentes que entraram num ensaio devem ser referenciados na sua conclusão.

Se isto não acontece, ou se números substanciais de doentes são perdidos para o *follow-up*, a validade do estudo é susceptível de ser posta em causa, sendo o grau de comprometimento dos resultados do estudo tanto maior, quanto maior for o número dos doentes perdidos. Este facto deve-se a que muitas vezes os doentes perdidos têm prognósticos diferentes dos que se mantêm no estudo. Os doentes podem desaparecer porque sofreram resultados adversos, ou porque estão a evoluir bem e não compareceram nas consultas agendadas. A situação é a mesma que justifica uma análise de tipo intenção-de-tratar: os doentes que suspendem a sua medicação podem ter menos ou, em regra, mais probabilidade de sofrer o evento adverso alvo<sup>(24)</sup>.

Ao ler um artigo, cada clínico pode verificar por si próprio se a perda para o *follow-up* foi excessiva, assumindo, num ensaio positivo, que todos os doentes perdidos do grupo tratamento tiveram um resultado adverso, e todos os perdidos do grupo controle tiveram um bom resultado (cenário da pior hipótese) e calculando de novo os resultados na base destas premissas<sup>(1, 3)</sup>: se as conclusões do ensaio não mudam, então a perda para o *follow-up* não foi excessiva. Se as conclusões mudam, a força da inferência é enfraquecida, dependendo o grau deste enfraquecimento da probabilidade de que os doentes do grupo tratamento perdidos para o *follow-up* tenham um mau resultado, enquanto os doentes do grupo controle também perdidos tenham um bom resultado. Deste modo, o grau de comprometimento da validade do estudo permanece uma questão de julgamento clínico.

from the treatment group lost to follow-up having an adverse outcome, while the patients lost from the control group had a good outcome. Thus, the degree to which the validity of the trial is compromised remains a question of clinical judgement.

Pedido de separatas para:

Address for reprints:

ANTÓNIO VAZ CARNEIRO

Centro de Estudos de Medicina Baseada na Evidência

Piso 6

Av. Prof. Egas Moniz

1649-028 LISBOA

## BIBLIOGRAFIA / REFERENCES

1. User's Guides to the Medical Literature. A Manual for Evidence-Based Clinical Practice. Chicago: JAMA & Archives Journals AMA 2001.
2. Moore TJ. *Deadly Medicine*. New York: Simon & Schuster 1995.
3. Greenhalgh T. *How to read a paper. The basics of evidence-based medicine*. London: BMJ Publishing Group 2001.
4. Brennan P, Croft P. Interpreting the results of observational research: chance is not such a fine thing. *BMJ* 1994; 309:727-30.
5. *Qualitative Research in Health Care*. London: BMJ Publishing 2000.
6. Jadad AR. *Randomised Controlled Trials*. London: BMJ Books 1998.
7. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000;342:1878-86.
8. Guyatt GH, Sackett DL, Cook DJ, The Evidence-Based Medicine Working Group. Users' guides 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: 2598-601.
9. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-Based Medicine. How to practice and teach EBM*. Edinburgh: Churchill Livingstone 2000.
10. Crombie IK. *The pocket guide to critical appraisal*. London: BMJ Publishing Group 1996.
11. Roberts C, Torgerson D. Randomisation methods in controlled trials. *BMJ* 1998;317:1301.
12. Roberts C, Torgerson D. Baseline imbalance in randomised controlled trials. *BMJ* 1999;319:185.
13. Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. *Lancet* 2001;357:373-80.
14. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. *JAMA* 1995;273:408-12.
15. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology*. Boston: Little, Brown and Company 1991.
16. Horwitz RI, Viscoli CM, Berkman L, et al. Treatment adherence and risk of death after a myocardial infarction. *Lancet* 1990;336:425-545.
17. Dawes M, Davies P, Gray A, Mant J, Seers K, Snowball R. *Evidence-based practice*. Churchill Livingstone 1999.
18. Elwood M. *Critical appraisal of epidemiological studies and clinical trials*. Oxford: Oxford University Press 1998.
19. Kaptchuk TJ. Powerful placebo: the dark side of the randomised controlled trial. *Lancet* 1998;351:1722-5.
20. Laporte JR, Figueras A. Placebo effects in psychiatry. *Lancet* 1994;344:1206-9.
21. Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *N Engl J Med* 2001;344:1594-602.
22. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;317:1185-90.
23. Guyatt GH, Pugsley SO, Sullivan MJ, et al. Effect of encouragement on walking test performance. *Thorax* 1984;39: 818-22.
24. Montori VM, Guyatt GH. Intention-to-treat analysis. *Can Med Assoc J* 2001;165:1339-41.