

AVALIAÇÃO CRÍTICA DE ARTIGOS RELATIVOS A ENSAIOS CLÍNICOS CONTROLADOS

II. Resultados

J.J. FERREIRA, C. SAMPAIO
Instituto de Farmacologia e Terapêutica Geral.
Faculdade de Medicina de Lisboa. Lisboa.

RESUMO

A forma de apresentação dos resultados de um ensaio clínico influencia a atitude dos médicos. São apresentados métodos e parâmetros para análise e valorização clínica dos resultados dos ensaios clínicos (risco relativo, redução do risco absoluto, número de doentes que é necessário tratar, intervalos de confiança, etc.).

SUMMARY

Critical Assessment of Articles on Controlled Clinical Trials II. Results

The way trial results are reported affects the physician's decisions. We describe some methods for the critical appraisal of clinical trial results (relative risk, absolute risk reduction, number needed to treat, confidence intervals, etc.).

INTRODUÇÃO

Quando se lê um artigo científico, nomeadamente um ensaio clínico, vários são os aspectos que potencialmente influenciaram as atitudes do leitor. Em relação a alguns desses factores o leitor não tem qualquer influência e na melhor das hipóteses reconhece a possibilidade de estar a ser influenciado nas suas conclusões. São exemplo deste aspecto o enviesamento resultante da simples publicação de um ensaio clínico, facto este designado por *publication bias*¹⁻⁴. A submissão preferencial pelos investigadores de ensaios com resultados *positivos*¹⁻³, bem como a maior aceitação desses mesmos ensaios por

parte dos editores e revisores das revistas médicas¹⁻³ vão, por um lado, influenciar as conclusões retiradas pela comunidade científica, e por outro, dificultar a realização de meta-análises e revisões sistemáticas, fundamentais para produzir, o mais precocemente possível, a evidência científica que suporta a utilização, ou não, de uma determinada intervenção terapêutica.

O enviesamento das publicações na direcção dos resultados positivos, embora humanamente compreensível, é cada vez menos legítimo devido à dificuldade em produzir um resultado final cumulativo de todos os dados, mas também por poder ser considerado má prática cien-

tífica⁵ a não submissão para publicação de resultados *negativos*. Estes estudos envolvem a participação voluntária de doentes que, como *consumidores*, têm direito a possuir toda a informação.

No extremo oposto encontramos o momento individual da decisão clínica. Independentemente dos argumentos teóricos disponíveis, a atitude terapêutica é influenciada por factores individuais relacionados com o doente e com o médico. É impossível controlar num ensaio todos os eventuais factores de prognóstico do doente (ex. capacidade económica, nível educacional, apoio familiar, adaptação no emprego, etc.). O mesmo sucede com a atitude médica, que incorpora processos não verbais que influenciarão a interacção médico-doente e consequentemente o acto terapêutico.

Entre estes dois momentos, encontramos a fase em que o clínico tem acesso e interpreta a informação científica disponível.

Na análise individual dos resultados de um ensaio clínico publicado várias são as questões a colocar^{6,7}:

- 1) Quais os *verdadeiros* resultados?
- 2) Qual o significado clínico desses resultados?
- 3) Podem os resultados apresentados ser extrapolados para o *meu* doente?

AVALIAÇÃO CRÍTICA DOS RESULTADOS DE UM ENSAIO CLÍNICO PUBLICADO

1. Quais os verdadeiros resultados?

O primeiro factor a ter em consideração é o tipo de variáveis utilizadas no estudo. As variáveis podem ser categoriais (ex. mortalidade, hospitalização, dor, etc.) em que assumem sempre um de dois resultados possíveis, ou variáveis contínuas (ex. duração, peso, valor de tensão arterial, etc.). Na sua análise estatística os princípios analíticos são semelhantes, mas existem ainda grandes limitações metodológicas na análise das variáveis contínuas. Por uma questão de simplicidade os exemplos referidos irão considerar unicamente variáveis categoriais.

O outro aspecto a ter em análise é o significado clínico das variáveis estudadas. As variáveis laboratoriais (ex. HDL, CD4+) ou imagiológicas (ex. placas de desmielinização na esclerose múltipla) são habitualmente mais facilmente quantificáveis do que as variáveis clínicas (ex. incapacidade, função, morbidade), sendo designadas por *surrogate end points* (variável que pode estar correlacionada com um efeito clínico mas que não é uma medida directa desse efeito). Contudo, não estando demonstrada uma estreita correlação entre os

valores obtidos num exame complementar e a repercussão no estado clínico, estes devem ser cautelosamente interpretados. Um exemplo clássico são os ensaios clínicos desenhados para provar a eficácia de novos fármacos na osteoporose na mulher após a menopausa, em que são usadas como variáveis primárias os valores da densitometria óssea em detrimento da variável clínica *ocorrência de fractura óssea*.

Está também demonstrado que a forma numérica como os resultados são apresentados influi na interpretação dos mesmos e consequentemente na atitude dos médicos aos quais é fornecida essa informação⁹⁻¹¹.

Os resultados devem ser apresentados de uma forma clara, facilmente interpretáveis e descrevendo os valores referentes a cada um dos participantes, de forma a poderem ser também trabalhados pelo leitor ou pelo investigador que decida realizar uma meta-análise ou revisão sistemática. A apresentação dos dados individuais de cada participante, possibilita ainda a realização de revisões sistemáticas com base na análise individual dos dados de cada doente, factor que diminui ainda mais o enfiamento¹².

De uma forma simples, a partir da leitura do artigo, deve ser possível obter, para cada variável analisada, os valores que possibilitem o preenchimento do Quadro I (Quadro I). O termo *evento* é apresentado como sinónimo de um acontecimento clínico em estudo.

Quadro I

	Resultados outcomes		Total
	Ocorrência de um evento	Não ocorrência de um evento	
Grupo controle (placebo)	a	b	a+b
Grupo experimental (intervenção em estudo)	c	d	c+d

Conhecendo os valores a, b, c e d é possível obter, através de simples cálculos, diferentes formas de apresentação dos resultados (Quadro II)^{6,13,14}.

Como exemplo da importância que diferentes formas de apresentação de resultados têm na conclusão a retirar de um artigo, e mais importante ainda na decisão clínica consequente, apresentam-se os resultados do estudo *Helsinki Heart Study: primary-prevention trial with gemfibrozil in middle-aged men with dyslipidemia*, publicado em 1987 no *The New England Journal of Medicine*¹⁵. É um estudo aleatorizado, em dupla ocultação, com a duração de cinco anos, comparativo do gemfibrozil com placebo, em homens de meia idade com dislipidemia, e cujo objectivo era avaliar a incidência de doença coronária. Na descrição dos resultados são apre-

Quadro II

	Fórmula de cálculo	Conceito
Risco absoluto no grupo controle (RAC)	$a/a+b$	Probabilidade de ocorrência de um evento no grupo controle
Risco absoluto no grupo experimental (RAE)	$c/c+d$	Probabilidade de ocorrência de um evento no grupo experimental
"Odds" no grupo controle	a/b	Probabilidade de ocorrência de um evento <i>contra</i> a probabilidade de não ocorrer, no grupo controle.
"Odds" no grupo experimental	c/d	Probabilidade de ocorrência de um evento <i>contra</i> a probabilidade de não ocorrer, no grupo experimental
Risco relativo (RR)	RAE/RAC	Probabilidade relativa de ocorrer um evento no grupo experimental quando comparado com o grupo controle
Redução do risco relativo (RRR)	$100\%-RR$	Valor de redução do risco de ocorrência de um evento no grupo experimental comparativamente com o grupo controle
"Odds ratio"	"Odds" no grupo exp. "Odds" no grupo controle	Relação entre o valor de <i>odds</i> no grupo experimental e controle; este valor aproxima-se do RR quando a incidência do evento é pequena
Redução do risco absoluto (RRA)	$RAC-RAE$	Valor de redução do risco absoluto no grupo experimental comparativamente com o grupo controle
Número necessário tratar (NNT)	$1/RRA$	Número de doentes que é necessário tratar para prevenir um evento, num determinado período de tempo

sentados os dados referentes aos doentes que terminaram o estudo e aos doentes que abandonando o estudo continuaram a ser acompanhados, sendo assim possível realizar uma análise em intenção de tratar. Assim, nessa análise os doentes que abandonaram o estudo foram analisados no grupo para onde foram aleatorizados inicialmente. É assim possível obter os valores de *a* a *d* indispensáveis para o preenchimento do Quadro I (Quadro III).

Quadro III

	Acidentes cardiovasculares em doentes que completaram o estudo	Acidentes cardiovasculares em doentes que abandonaram o estudo	Total de acidentes cardiovasculares	Total de doentes sem ocorrência de acidentes cardiovasculares	Total
Grupo placebo	72	12	a 84	b 1946	a+b 2030
Grupo tratado com gemfibrozil	46	10	c 56	d 1995	c+d 2051

Com os valores presentes no Quadro III é possível calcularmos vários parâmetros de apresentação dos resultados (Quadro IV).

Quadro IV

	Resultado
Risco absoluto no grupo placebo (RAC)	4,1%
Risco absoluto no grupo experimental (RAE)	2,7%
"Odds" no grupo controle	4,3%
"Odds" no grupo experimental	2,8%
Risco relativo (RR)	66%
Redução do risco relativo (RRR)	34%
"Odds ratio"	65%
Redução do risco absoluto (RRA)	1,4%
Número necessário tratar (NNT)	72 (num período de 5 anos)

Temos assim um ensaio clínico, publicado numa das revistas médicas de maior impacto e com vários valores de resultados correspondentes a diferentes formas de apresentação. Curiosamente, no artigo original, o resultado é apresentado unicamente em redução do risco relativo (RRR) que é de 34%. Numa leitura imediata, é fácil concluir pelo maior impacto de apresentação dos resultados em RRR do que em Redução do Risco Absoluto (RRA) (1,4%). Este estudo teve a particularidade de ter sido usado para investigar a repercussão da forma de apresentação dos seus resultados no acto médico^{9,10}. Num dos estudos⁹, foi distribuído a 148 clínicos gerais os resultados deste estudo, sob a forma de redução do risco relativo, redução do risco absoluto, diferenças no número de doentes sem acidentes cardiovasculares, número de doentes que era necessário tratar e redução de acidentes e mortalidade, como se se tratasse de resultados de diferentes estudos. Foi-lhes solicitado perante cada resultado que indicassem numa escala analógica a sua vontade de prescrever aquele fármaco. Como resultado foi obtida um concordância entre os médicos no sentido da prescrição do fármaco em 77% dos casos quando a forma de apresentação era a redução do risco relativo e unicamente de 24% quando apresentada em redução do risco absoluto. Curiosamente nenhum dos clínicos se apercebeu que todos os valores correspondiam aos mesmos resultados.

2. Qual é o significado clínico dos resultados?

Se em relação ao estudo já referido fizermos variar o risco absoluto de ocorrência da variável em estudo, tornado-a dez vezes superior, iremos obter diferentes resultados (Quadro V).

O facto óbvio de uma maior incidência do evento que se está a medir ter repercussão no significado atribuído aos resultados de um ensaio, só tem tradução quando estes são apresentados na forma de RRA ou NNT, dado

Quadro V

	RAC 4,1% RAE 2,7%	RAC x 10 = 41% RAE x 10 = 27%
Redução do risco relativo (RRR)	34%	34%
Odds ratio	65%	65%
Redução do risco absoluto (RRA)	1,4%	14%
Número necessário tratar (NNT)	72	8

que a RRR e o Odds ratio, pela forma como são calculados, não têm em linha de conta a frequência de ocorrência do evento. Podemos chegar à situação extrema de termos uma intervenção terapêutica com uma eficácia traduzida em termos de RRR num valor de quase 100%, mas que na prática não tem qualquer relevância clínica dada a raridade do evento que se pretende evitar.

Sabendo as limitações e o tipo de informação fornecido por cada um dos parâmetros, qual deve ser o escolhido para apresentar resultados?

O parâmetro Número Necessário Tratar (NNT) é indiscutivelmente aquele cuja leitura mais directamente fornece a informação que o clínico requer para tomar uma decisão. Para além de ser facilmente interpretável, incorpora não só a componente eficácia comparativa da intervenção em estudo mas também a incidência do evento em avaliação. O problema é colocado mais uma vez na valorização do valor obtido. A primeira regra é de bom senso, pelo que um valor de oito (Quadro V) tem seguramente um importante interesse clínico. Em relação ao valor 71 a sua interpretação deve ser feita comparando com outros NNT já conhecidos e aplicados na prática clínica corrente¹⁶ (Quadro VI). Na leitura de um valor de NNT deve ser tido em consideração que este valor é referido sempre a um intervalo de tempo definido.

Por fim, devem ser conhecidos os efeitos adversos da mesma intervenção e é igualmente possível calcular o número de doentes que é necessário tratar, num determinado intervalo de tempo para causar um efeito adverso,

Quadro VI

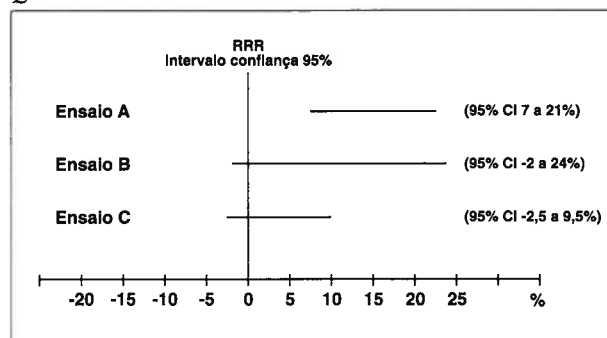
Situação clínica	Intervenção terapêutica	Eventos a serem prevenidos	Duração do seguimento	NNT
Trabalhadores de saúde ¹⁷	Vacina contra hepatite B	Hepatite B	1 ano	7 (risco base 200 casos/1000/ano -145 (risco base 10/1000/ano)
Idoso (≥ 60 anos) ¹⁸	Fármacos antihipertensores	Mortalidade total	5 anos	12
Eclâmpsia	Sulfato de Mg e.v. (vs diazepam)	Recorrência de convulsões	horas	7
Recém-nascidos pré-termo ²⁰	Corticosteróides pré-parto	Síndrome de dificuldade respiratória do RN	dias	11

sendo designado por NNH (*number needed to harm*).

Os resultados da análise estatística de um estudo comparativo são habitualmente apresentados em termos de significância da diferença entre dois ou mais grupos²¹. É previamente definido um valor de p (correspondente ao erro tipo a admitido) abaixo do qual é considerada a diferença como significativa. Assim, admitir como provada uma diferença quando $p < 0,05$ é assumir que o achado dessa diferença pode ser devido ao acaso em apenas uma em cada cinco vezes. Por outro lado, se o valor de p for maior que 0,05 diremos que as duas intervenções não são diferentes ou que a dimensão da amostra foi insuficiente para provar a diferença. Obtemos assim conclusões taxativas, pouco plásticas e conseqüentemente distantes dos mecanismos de raciocínio clínico presentes na tomada de decisões. Um resultado estatisticamente significativo pode, por outro lado, não ter qualquer relevância clínica.

Em contraponto com esta abordagem dicotômica, surge a análise comparativa de duas intervenções em que se apresenta para um qualquer dos testes estatísticos ou parâmetros utilizados (*t student*, RRR, odds ratio, NNT) um intervalo de confiança em redor do valor obtido^{13,22,23}. Este intervalo de confiança significa que, quando calculamos, por exemplo, um valor de redução do risco relativo, o valor *real* da diferença fica dentro dos limites do intervalo de confiança em 95% dos casos (se calculado o intervalo de confiança para o valor 95%). Assim, estamos a assumir que se repetirmos inúmeras vezes o mesmo ensaio, em 95% dos casos o valor encontrado para o parâmetro em estudo cai dentro dos limites do intervalo de confiança. Esta forma de apresentação permite-nos avaliar de uma forma mais realista o verdadeiro valor interpretativo dos resultados obtidos. Nos exemplos apresentados no Quadro VII, no ensaio A o intervalo de confiança não inclui o valor zero de RRR (diferença nula), pelo que numa abordagem dicotômica diríamos de imediato que a diferença era sempre favorável ao tratamento em estudo (Quadro VII). A van-

Quadro VII



tagem de apresentar os resultados com intervalos de confiança coloca-se fundamentalmente quando os ensaios são *negativos* (ensaios B e C). O ensaio B, em que o intervalo de confiança inclui o valor 0 de RRR, pode corresponder unicamente ao ensaio A, mas em que foi usada uma amostra de menor dimensão, alargando assim os intervalos de confiança. Num estudo *negativo* deve também ser considerado se esta conclusão é definitiva¹³. No ensaio C, se considerarmos o valor mais alto do intervalo de confiança (valor de máximo benefício) existe unicamente 2,5% de probabilidade do valor real da diferença ser igual ou superior a este valor. Se mesmo este valor de possível máximo benefício não tiver qualquer relevância clínica, muito provavelmente poderemos assumir esta ausência de diferença com um carácter quase definitivo.

3. Os resultados podem ser aplicados ao meu doente?

Na situação óptima de o doente em questão corresponder aos critérios de inclusão dos estudos que evidenciaram o benefício da terapêutica, a decisão é lógica e imediata.

O problema surge quando o nosso doente difere numa ou em mais características dos doentes dos estudos. O que é sugerido pelos metodologistas é tentar averiguar se as propriedades em que difere o nosso doente dos doentes *ideais* são aspectos demonstrados como factores de prognóstico na evolução da sua situação clínica⁶. Se não houver evidência que indique que o doente venha obrigatoriamente a apresentar um comportamento diferente, muito provavelmente os resultados poderão ser extrapolados para o doente em causa⁶.

A decisão final perante o doente permanecerá sempre uma decisão pontual, no sentido em que incorporará sempre critérios de decisão quase impossíveis de estudar de forma sistemática e que se traduzem por preferências individuais (ex. critérios religiosos, condições económicas do indivíduo, incapacidade funcional, actividade profissional), mas este facto não deve servir para impedir que o clínico, independentemente da decisão final, possua um conhecimento válido e actualizado de qual⁸ a evidência científica sobre a eficácia e tolerabilidade da intervenção em questão.

CONCLUSÃO

Vários são os factores relacionados com a publicação de um ensaio clínico que influem nas atitudes dos médicos que têm acesso a essa fonte de informação. Alguns aspectos escapam ao controlo dos *leitores*, enquanto ou-

tros poderão ser por estes analisados de forma crítica.

A análise dos resultados de um ensaio clínico requer o uso de métodos específicos e para a sua correcta interpretação é fundamental conhecer alguns dos seus aspectos básicos, nomeadamente as limitações resultantes da sua interpretação linear.

A informação fornecida pelos ensaios clínicos individuais deve ser assim correctamente interpretada, contextualizada de forma a que a decisão final no tratamento do *nosso doente* tenha um claro suporte científico.

BIBLIOGRAFIA

- DICKERSIN K: How important is publication bias? *AIDS Educ Prev* 1997; 9 Suppl A: 15-21.
- DICKERSIN K, MIN YI, MEINERT CL: Factors influencing publication of research results. Follow up of applications submitted to two institutional review boards. *JAMA* 1992; 263: 374-8.
- DICKERSIN K: The existence of publication bias and risk factors for its occurrence. *JAMA* 1990; 263: 1385-9.
- DICKERSIN K, CHAN S, CHALMERS TC, SACKS HS, SMITH H JR: Publication bias and clinical trials. *Controlled Clin Trials* 1987; 8:343-53.
- CHALMERS TC, FRANK CS, REITMAN D: Minimising the three stages of publication bias. *JAMA* 1990; 263: 1392-5.
- GUYATT GH, SACKETT DL, COOK DJ: User's guide 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? *JAMA* 1994; 271: N°1.
- RICHARDSON WS, DETSKY AS: User's guides to the medical literature. VII. How to use a clinical decision analysis. A. Are the results of the study valid? *JAMA* 1995; 273: N°16.
- GREENHALGH T: Papers that report drug trials. *BMJ* 1997; N° 7106; Vol 315.
- BOBBIO M, DEMICHELIS B, GIUSTETTO G: Completeness of reporting trial results: effect on physicians' willingness to prescribe. *Lancet* 1994; 343: 1209-11.
- NAYLOR CD, CHEN E, STRAUSS B: Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Annals of Internal Medicine* 1992; 117:916-21.
- BUCHER HC, WEINBACHER M, GYR K: Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. *BMJ* 1994; 309:761-4.
- STEWART LA, PARMAR MK: Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet* 1993; 341:418-22.
- GREENHALGH T: Statistics for the non-statistician. II: "Significant" relations and their pitfalls. *BMJ* 1997; N°7105; Vol 315.
- SINCLAIR JC, BRACKEN MB: Clinically useful measures of effect in binary analyses of randomized trials. *J Clin Epidemiol* 1994; 47; N°8: 881-9.
- FRICK MH, ELO O, HAAPA K et al.: Helsinki heart study: primary-prevention trial with gemfibrozil in middle-aged men with dyslipidemia. *N Engl J Med* 1987; 317:1237-45.
- SACKETT DL, RICHARDSON WS, ROSENBERG W, HAYNES RB: Evidence-based medicine. How to practice and teach EBM. Churchill Livingstone 1997.
- JEFFERSON T, DEMICHELI V, DEEKS J, MACMILLAN A, SASSI F, PRATT M: Vaccines against hepatitis B in health-care workers. In: Gluud C, Jørgensen T, Koretz RL, Morabito A, Pagliaro L, Poynard T, Sutton R (eds.) *Hepato-Biliary Module of The Cochrane Database of Systematic Reviews*, [updated 02 December 1997]. Available in The Cochrane Library [database on disk and CDROM]. The Cochrane Collaboration; Issue 1. Oxford: Update Software 1998.

Updated quarterly.

18. MULROW C, LAU J, CORNELL J, BRAND M: Antihypertensive drug therapy in the elderly. In: GUEYFFIER F, LAU J, MULROW CD (eds.): Hypertension Module of The Cochrane Database of Systematic Reviews, [updated 02 December 1997]. Available in The Cochrane Library [database on disk and CDROM]. The Cochrane Collaboration; Issue 1. Oxford: Update Software 1998. Updated quarterly.

19. DULEY L, HENDERSON-SMART D: Magnesium sulphate versus diazepam for eclampsia. In: Neilson JP, Crowther CA, Hodnett ED, Hofmeyr GJ (eds.): Pregnancy and Childbirth Module of The Cochrane Database of Systematic Reviews, [updated 02 December 1997]. Available in The Cochrane Library [database on disk and CDROM]. The Cochrane Collaboration; Issue 1. Oxford: Update Software 1998. Updated quarterly.

20. Crowley P: Corticosteroids prior to preterm delivery. In: Neilson JP, Crowther CA, Hodnett ED, Hofmeyr GJ (eds.): Pregnancy and Childbirth Module of The Cochrane Database of Systematic Reviews, [updated 02 December 1997]. Available in The Cochrane Library [database on disk and CDROM]. The Cochrane Collaboration; Issue 1. Oxford: Update Software 1998. Updated quarterly.

21. GUYATTT G, JAENSCHKE R, HEDDLE N, COOK D, SHANNON H, WALTER S: Basic statistics for clinicians. 1. Hypothesis testing. *Can Med Assoc J* 1995;152:27:32.

22. WALTER SD: Methods of reporting statistical results from medical research studies. *Am J Epidemiol* 1995; 141: 896-906.

23. GOODMAN SN, BERLIN JA: The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med* 1994; 121:200-6.