

ENSAIOS CLÍNICOS À LUZ DA ESTATÍSTICA CLINICAL TRIALS UNDER THE LIGHT OF STATISTICS

Reinaldo José Gianini*⁵

“é preciso com frequência muita fé para poder generalizar os resultados de ensaios clínicos à prática usual.”
FletcherRH

Os ensaios clínicos randomizados são considerados padrão-ouro para a produção das evidências referentes à eficácia dos tratamentos. Ensaios clínicos de boa qualidade incorporam análises estatísticas por vezes muito complexas e que resultam em dados de difícil compreensão por parte dos médicos, justamente os indivíduos mais interessados na questão.

Há vários pontos a serem analisados para a boa compreensão dos ensaios clínicos, tais como erros aleatórios, erros sistemáticos, validade externa, análise estatística, desfecho estudado, concepção teórica da rede causal, técnica de amostragem, etc. O objetivo do presente trabalho é destacar alguns pontos cruciais de interesse neste tema, debater e contribuir para o entendimento de seus significados e implicações, especificamente: como julgar uma medida de significância, o valor do p , ao lado de outras medidas de associação, de efeito e de impacto, de modo a obter a melhor decisão sobre a relevância dos resultados de um ensaio clínico.

O VALOR DO P

Atualmente, em alguns círculos, ainda observamos a deificação do valor do p estatístico, como se o fato de ser significativo ou não decidisse definitivamente a questão sobre a eficácia de determinado tratamento. Nesse sentido, é importante resgatar seu significado.

Em toda e qualquer pesquisa, por razões relacionadas a recursos e à ética, não se trabalha com o universo populacional, mas sim com amostras. Assim, os ensaios clínicos na fase III, por exemplo, têm a preocupação de incluir na amostra um número mínimo suficiente de pacientes para comprovar sua hipótese.

É aqui que reside um primeiro problema: por melhor que seja a técnica utilizada na amostragem, sempre há chance da amostra não representar idealmente a população. Em outras palavras, os resultados encontrados na amostra poderiam não ser os mesmos encontrados na população. A esta chance se dá o nome de erro aleatório quando depende do acaso e de variáveis não controláveis por parte do pesquisador.

O p estatístico descrito nos estudos se refere ao erro (aleatório) tipo I, denominado α , que indica a significância dos resultados encontrados na amostra: a chance da diferença entre grupo de tratamento e grupo controle (placebo) ter ocorrido ao acaso. Por convenção, nos estudos da área de Ciências Biológicas, incluindo a Medicina, o limite tem sido 0,05. Uma diferença no resultado entre grupo de tratamento e

grupo controle com p igual a 0,02 é considerada significativa porque implica em assumir que a chance deste resultado ter ocorrido ao acaso é de apenas duas em 100, tendo-se, como complemento, 98 chances em 100 de ser atribuído realmente ao tratamento.

No entanto, para um mesmo valor de p , significativa para a diferença entre grupos, a relevância dos resultados pode variar, como pode variar também a precisão dos resultados.

Por outro lado, um valor de p não significativa pode ocorrer mesmo com resultados relevantes. Neste caso, porém, muito imprecisos. E isto ocorre em amostras de tamanho insuficiente, cujo poder estatístico ($1-\beta$) ficou aquém do desejado. β é o erro (aleatório) tipo II, que significa rejeitar a diferença entre grupos quando esta for verdadeira. Normalmente se trabalha com um poder estatístico igual a 0,80 (portanto, β igual a 0,20).

Analisando-se os exemplos da tabela 1, que propositadamente reúne estudos com diferentes tamanhos de amostra e diferentes resultados, tem-se: os estudos 1, 2 e 3 apresentam o mesmo valor de p , no entanto, apresentam efeitos de diferentes magnitudes.

O estudo 3 é o que apresenta maior risco relativo (RR), maior redução absoluta do risco (RAR), maior redução relativa do risco (RRR) e menor número necessário para tratar (NNT), sendo todos muito relevantes. Entretanto, tem um poder estatístico insatisfatório e um intervalo de confiança (IC95%) muito impreciso, explicado principalmente por seu pequeno tamanho amostral.

O estudo 2 apresenta o menor RR, a menor RAR, a menor RRR e o maior NNT, que poderiam ser julgados de pouca relevância. No entanto, tem poder estatístico satisfatório e um intervalo de confiança de maior precisão. Mas o limite inferior de seu intervalo de confiança está muito próximo de 1, o que indicaria ausência de diferenças de risco entre o grupo controle e grupo de tratamento.

O estudo 1 apresenta RR, RAR, RRR e NNT intermediários, que podem ser considerados relevantes, e poder estatístico bastante satisfatório, apesar de um IC95% de precisão intermediária.

Considerando-se todas estas informações, e não apenas o valor do p provavelmente ficaríamos mais seguros em adotar, escolhendo-se entre estes três tratamentos, o tratamento pesquisado pelo estudo 1. Descartaríamos o tratamento do estudo 3 e acharíamos de pouca vantagem o tratamento do estudo 2.

Observe que, mais importante que adotar ou descartar, aumentamos consideravelmente nosso poder de crítica, e isto é fundamental.

⁵vide definições das medidas de associação, de efeito e de impacto no Quadro I.

Agora vejamos os exemplos com valor de p não significante (Tabela 2): provavelmente, se tivéssemos que escolher entre estes três tratamentos, descartaríamos os tratamentos dos estudos 5 e 6 com certa tranquilidade, dada a irrelevância das diferenças. Mas, certamente, ficaríamos inseguros em simplesmente descartar o tratamento do estudo 4, porque os resultados são muito relevantes, e talvez numa amostra de maior tamanho (com poder estatístico satisfatório), revelaria diferenças significantes.

De novo, enfatize-se, conseguimos fazer uma análise mais completa.

DECISÃO CLÍNICA E DECISÃO DE GOVERNO

Uma medida de impacto importante a ser observada é o risco atribuível na população (RAP). Levando em consideração a prevalência da doença na população, esta medida possibilita a estimação do número de eventos que poderiam ser evitados se adotado um determinado tratamento. É mais um indicador de relevância.

Tomemos como exemplos os estudos 1 e 2 (Tabela 1). Digamos que a doença pesquisada no estudo 1 tenha uma baixa prevalência (1%) e a doença pesquisada no estudo 2 tenha uma prevalência maior (30%). Numa população de um milhão de habitantes, o RAP do tratamento pesquisado no estudo 1 seria 1.200 óbitos evitados ($= 0,12 \times 0,01 \times 1.000.000$), e do estudo 2 seria 18.900 óbitos evitados ($= 0,063 \times 0,3 \times 1.000.000$). Portanto, mesmo o estudo 2 tendo

RR e RAR menores, seu tratamento acaba sendo uma intervenção populacional de maior impacto porque a prevalência da doença de que trata é bem maior.

Outro aspecto a se considerar é o custo-benefício: supondo-se, primeiramente, que o estudo 1 tenha sido sobre uma doença rara na população, com prevalência de dez casos para um milhão de habitantes, e que cada tratamento custe R\$ 5.000,00 (situação A). Teríamos que despende R\$ 41.500,00 para se evitar um óbito ($= 8,3 \times R\$5.000,00$). Supondo-se agora que o mesmo estudo tenha sido sobre uma doença de elevada prevalência, igual a 20%, ou seja, 200.000 casos em um milhão, e que cada tratamento custe R\$ 200,00 (situação B). Com R\$ 41.500,00 poderíamos tratar 208 casos e evitar 25 óbitos ($= 208/8,3$), ou R\$ 1.660,00 ($= R\$41.500,00/25$) por óbito evitado. Nesta última situação, se todos os doentes nessa população fossem tratados, seriam evitados 24.096 óbitos ($= 200.000/8,3$).

Decisões clínicas são tomadas colocando-se o interesse do paciente em primeiro lugar. Decisões de governo são tomadas colocando-se o interesse coletivo em primeiro lugar. Assim, apesar de parecer justo ao clínico evitar o óbito de seu paciente a qualquer custo (situação A), pode-se compreender porque para determinados governos haveria interesse apenas em bancar o custo na situação B.

Em conclusão: não podemos tomar decisões sobre qual o melhor tratamento com base apenas no valor do p . É imprescindível analisar outros aspectos, principalmente o NNT e o IC95%.

Quadro I - Medidas de associação, de efeito e de impacto descritas nos ensaios clínicos*

MEDIDA	MÉTODO DE CÁLCULO	SIGNIFICADO
Risco relativo (RR)	Incidência do evento no grupo controle / Incidência do evento no grupo de tratamento	Razão de risco entre os grupos. Expressa quantas vezes o risco é maior no grupo controle.
Redução absoluta do risco (RAR)	Incidência do evento no grupo controle - incidência do evento no grupo de tratamento	Diferença de risco entre os grupos atribuída ao efeito do tratamento. Número de eventos evitados no grupo de tratamento.
Redução relativa do risco (RRR)	RAR / Incidência no grupo controle	Proporção do risco redutível pelo tratamento.
Número necessário para tratar (NNT)	1 / RAR	Quantidade de indivíduos que precisam ser tratados para se evitar um evento.
Intervalo de 95% de confiança (IC95%)	$1 \pm 1,96 \sqrt{\frac{X^2}{RR}}$	Límite dentro do qual há 95% de chance de se situar o verdadeiro RR (da população).
Risco atribuível na população (RAP)	RAR X prevalência da doença na população X número de indivíduos na população	Número de eventos evitados na população se todos os doentes forem tratados.

* X^2 : qui-quadrado

Tabela 1. Diferentes exemplos (fictícios) de ensaios clínicos* com resultados estatísticos significantes

Estudo	Valor do P	Poder (1-β)	Incidência no Grupo Controle	Incidência no Grupo Tratado	RR	IC95%	N	RAR	RRR	NNT
1	0,02	0,97	22 / 100	10 / 100	2,2	1,1 — 4,4	200	12/100	54,5%	8,3
2	0,02	0,81	49 / 300	30 / 300	1,6	1,07 — 2,5	600	6,3/100	38,3%	15,9
3	0,02	0,64	8 / 10	2 / 10	4,0	1,1 — 14,3	20	60/100	75,0%	1,7

* Evento analisado: óbito ao final do período de seguimento

Tabela 2. Diferentes exemplos (fictícios) de ensaios clínicos* com resultados estatísticos não-significantes

Estudo	Valor do P	Poder (1-β)	Incidência no Grupo Controle	Incidência no Grupo Tratado	RR	IC95%	N	RAR	RRR	NNT
4	0,07	0,42	8 / 10	3 / 10	2,7	0,98 — 7,22	20	50 / 100	62,5%	2
5	0,07	0,61	45 / 300	30 / 300	1,5	0,97 — 2,31	600	5 / 100	33,3%	20
6	0,07	0,72	19 / 100	10 / 100	1,9	0,93 — 3,88	200	9 / 100	47,4%	11,1

* Evento analisado: óbito ao final do período de seguimento

BIBLIOGRAFIA

1. Fletcher RH, Fletcher SW, Wagner EH. Epidemiologia clínica. 3a ed. Porto Alegre: Artes Médicas; 1996.
2. Medronho RA. Epidemiologia. Rio de Janeiro: Atheneu; 2002.
3. Rothman JK. Modern epidemiology. Boston: Little, Brown; 1986.
4. Rouquayrol MZ. Epidemiologia e saúde. 6a ed. Rio de Janeiro: Medsi; 2003.
5. Vieira S. Introdução à bioestatística. 3a ed. Rio de Janeiro: Campus; 1991.