

# Medicina baseada em evidências e hipertensão arterial

## Medicine based on evidences and hypertension

Editor: Álvaro Avezum

### Cálculo do tamanho da amostra em estudos randomizados em hipertensão arterial sistêmica: grandes amostras ou metodologia adequada?

### Sample size calculations in randomised trials on systemic arterial hypertension: huge size or adequate methodology?

Hélio Penna Guimarães<sup>1</sup>, Lígia Nasi Laranjeira<sup>1</sup>, Álvaro Avezum<sup>2</sup>, Leopoldo Soares Piegas<sup>3</sup>

#### INTRODUÇÃO

No ambiente da pesquisa clínica ou experimental, é rotineira a máxima de que “o investigador sempre deve calcular corretamente o tamanho de amostra antes de iniciar suas pesquisas randomizadas e descrever os detalhes pertinentes de forma adequada em suas publicações”. Assim, os cálculos do tamanho de amostra para estudos randomizados parecem indispensáveis, e distintos métodos são descritos em artigos científicos e livros de bioestatística<sup>1,2</sup>.

No entanto, há questionamentos sobre se um estudo realmente pode ser considerado antiético quando não apresenta um cálculo adequado do tamanho da amostra. Dessa forma, discute-se também se os estudos de baixo poder estatístico podem ser aceitáveis, desde que assegurados o rigor metodológico, o controle de vieses e, portanto, a possibilidade de sua incorporação em metanálises<sup>3,4</sup>.

#### COMPONENTES DE CÁLCULO DO TAMANHO DE AMOSTRA

Nos cálculos que determinam o efeito de uma intervenção, deve-se detectar, por exemplo, a taxa de evento no grupo com tratamento e no grupo controle. O cálculo do tamanho da amostra para estudos randomizados com grupo controle habitualmente requer a cuidadosa consideração de alguns conceitos fundamentais<sup>1,2</sup>:

- **Erro tipo I ou alfa** – Os investigadores podem concluir que dois tratamentos diferem quando, de fato, são iguais. Esse erro considera o poder do estudo analisando a taxa de eventos no grupo controle e o efeito da intervenção

sobre o grupo com tratamento (ou a taxa de eventos no grupo sob tratamento). Assim, com resultados quantitativos e um teste estatístico adequado, os investigadores têm subsídios para supor diferenças entre o significado e a variação desses resultados. Esse tipo de erro afere a probabilidade de uma conclusão ser falso-positiva. Convencionalmente, o erro alfa é definido freqüentemente entre 0,05, representando uma possibilidade menor que 5% de haver uma conclusão falso-positiva;

- **Erro tipo II ou beta** – Os investigadores podem concluir que dois tratamentos não diferem quando, de fato, eles o fazem (conclusão falso-negativa). Convencionalmente, o erro beta é freqüentemente definido entre 0,20, representando uma possibilidade menor que 20% de haver uma conclusão falso-negativa. O poder do estudo deriva do erro beta;
- **Poder do estudo** – O complemento de 1-beta representa a probabilidade de se evitar uma conclusão falso-negativa. Assim, para  $\beta = 0,20$ , o poder do estudo seria de 0,80 (ou 80%), ou seja, um estudo com poder de 80% tem 80% de chance de detectar uma diferença entre dois tratamentos se houver uma diferença real nessa população.

A despeito da complexidade conceitual, o desconhecimento do real significado do erro alfa, do erro beta e do poder do estudo pode comprometer o adequado desenho metodológico, que transforma o convencional cálculo do tamanho da amostra num problema pragmático.

Recebido: 27/07/2006 Aceito: 29/08/2006

1 Médico da Divisão de Pesquisa do Instituto Dante Pazzanese de Cardiologia, em São Paulo – SP.

2 Diretor da Divisão de Pesquisa do Instituto Dante Pazzanese de Cardiologia.

3 Diretor Geral do Instituto Dante Pazzanese de Cardiologia.

Correspondência para: Dr. Hélio Penna Guimarães. Divisão de Pesquisa – Instituto Dante Pazzanese de Cardiologia. Avenida Dr. Dante Pazzanese, 500, 13º andar – 04012-909 – São Paulo – SP. E-mail: heliopg@yahoo.com.br

## CALCULANDO O TAMANHO DA AMOSTRA

Tradicionalmente, como ponto de partida, estimam-se as taxas reais do evento nos grupos sob tratamento e controle quando, na realidade, seria recomendável estimar a taxa do evento na população geral e, só então, “determinar” o efeito do tratamento avaliado. Como exemplo, projeta-se um estudo para o uso de um novo anti-hipertensivo e se estima uma taxa do evento de 10% no grupo controle. Na seqüência, estimam-se a mudança absoluta (por exemplo, redução absoluta de 3%), a mudança relativa (redução relativa de 30%) ou simplesmente a taxa de eventos de 7% no grupo em tratamento com a nova medicação. Partindo dessas suposições, calcula-se então, o tamanho da amostra<sup>5-7</sup>. Geralmente, os pesquisadores utilizam *softwares* de tamanho da amostra (preferencialmente com orientação de um estatístico) e por vezes desconhecem como atribuir os valores ou buscar uma fonte adequada de informação para o correto cálculo da amostra.

## EFEITO DE SELEÇÃO DO ERRO ALFA E DO PODER

As convenções de  $\alpha = 0,05$  e poder = 0,80 geralmente são adequadas para a estimativa correta do tamanho da amostra. Entretanto, há de se considerar algumas questões: por exemplo, se um anti-hipertensivo fosse seguramente eficaz na prevenção de eventos cardiovasculares maiores com poucos efeitos colaterais, num estudo clínico de fase III poder-se-ia ajustar o erro alfa mais baixo (0,01) para reduzir ainda mais quaisquer possibilidades de um achado falso-positivo. Em geral, considera-se a redução do poder abaixo de 0,80 em função do conhecimento limitado sobre a ausência de tratamento eficaz. No entanto, se o cenário de estudo testa o novo anti-hipertensivo associado a um anti-hipertensivo previamente seguro e de baixo custo, pouco dano poderia ser causado. Então, ajustar o erro alfa entre 0,10 pode realmente fazer sentido<sup>7,8</sup>.

As projeções distintas do erro alfa e do poder alteram diretamente os tamanhos da amostra: reduzir alfa e aumentar o poder tem como consequência o aumento da amostra. Assim, ao se reduzir alfa de 0,05 a 0,01, promove-se o aumento de aproximadamente 70% no tamanho, com poder = 0,50. Com alfa = 0,05, a elevação do poder de 0,50 para 0,80 promove o aumento de 2 vezes no tamanho experimental e de 0,50 a 0,99 quase 5 vezes. Nesse contexto, as escolhas de alfa e poder produzem tamanhos de amostra e custos distintos nos estudos.

## ESTIMATIVA DOS PARÂMETROS POPULACIONAIS

A estimativa da taxa de eventos num grupo controle pode levar a estimativas irreais, por exemplo, num cenário fictício, é necessário estimar um evento raro com o uso de um determinado

anti-hipertensivo numa população específica, e a estimativa oficial é de 40%. Os pesquisadores, no entanto, consideram essa estimativa muito elevada e sugerem então 12%, estimando afinal, de forma conservadora, em 6%. Porém, o grupo placebo do estudo randomizado mostrou 1,9%. A primeira estimativa foi desconsiderada em mais de 20 vezes, o que afetou claramente o cálculo do tamanho da amostra. Em geral, as publicações provêm estimativas de desfechos no grupo controle, incorporando, contudo, diferenças como locações desiguais, critérios de elegibilidade, desfechos e tratamentos nesse cálculo. A despeito disso, essa informação caracteriza o ponto de partida<sup>7,8</sup>.

Em um estudo sobre prevenção de acidente vascular cerebral (AVC) com o uso de um novo anti-hipertensivo, os dados demonstraram que 10% dos pacientes apresentaram AVC no grupo com tratamento, tornando o evento raro no grupo controle. A estimativa do tamanho do efeito desejado deverá refletir os efeitos clínicos, e o aspecto de saúde pública não deve faltar na análise estatística. Nesse cenário, pode ficar sólida a evidência entre a diferença óbvia do uso de um anti-hipertensivo para a prevenção de AVC que custe 10 reais e desencadeie poucos efeitos colaterais e um novo anti-hipertensivo que custe 80 reais e apresente mais efeitos colaterais, porém com o uso amplamente encorajado.

Esses parâmetros apresentados possuem relevante impacto no cálculo do tamanho da amostra: mantendo os dados assumidos no grupo controle como constantes e dividindo ao meio as necessidades do tamanho do efeito, será necessário um aumento de 4 vezes no tamanho da amostra. De forma similar, se dividirmos por quatro o tamanho do evento, seria necessário um aumento de 16 vezes no tamanho da amostra. Inversamente, teríamos o aumento do tamanho da amostra pelo inverso do quadrado da redução do tamanho do efeito (o que estatisticamente se denomina de “relação ao quadrado”). Nesse cenário, com o parâmetro inicial estimado de 10% no grupo controle e 6% no grupo sob tratamento e  $\alpha = 0,05$  e poder = 0,90, 965 participantes seriam necessários para cada grupo. Dividindo ao meio o tamanho do efeito e a estimativa do grupo de intervenção para 8%, seria necessário um aumento de 4 vezes no tamanho da amostra para 4.301. Dividindo por quatro o tamanho do efeito por meio da alteração da estimativa do grupo de intervenção para 9%, seria necessário um aumento maior que 28 vezes no tamanho da amostra, 18.066 por grupo, permitindo a conclusão de que pequenas variações no tamanho do efeito geram enormes alterações no tamanho da amostra.

As necessidades de grandes amostras nos estudos com baixa taxa de evento frustram os investigadores. Essa frustração

decorre em parte da falta de compreensão em relação ao poder do estudo<sup>8-10</sup>.

## BAIXO PODER COM AVALIAÇÃO LIMITADA DE PARTICIPANTES

O que acontece quando o tamanho da amostra calculado no *software* (supondo a estimativa do investigador) produz um tamanho de mostra maior que o número de participantes avaliados? Frequentemente, os investigadores podem calcular e estimar o que possuem com baixo poder (0,40) para a avaliação dos participantes<sup>9</sup>.

O estudo de baixo poder pode ser aceitável se forem combinados em metanálise<sup>9,10</sup>. No entanto, três advertências devem ser consideradas:

1. *O estudo deve ser forte metodologicamente, eliminando viés* – Frequentemente, acredita-se que o poder adequado possa se sobrepor às discussões metodológicas. No entanto, a randomização inadequada pode provocar viés nos resultados que não podem ser salvos, ainda que em grandes amostras. Em contraste, se a metodologia é apropriada, o estudo resultará numa estimativa de efeito sem viés. E, mesmo com baixo poder (e precisão), o estudo poderia ser combinado com outros similares em uma metanálise;
2. *Os autores devem descrever seus métodos e resultados apropriadamente para evitar má interpretação* – Se os resultados do estudo forem descritos apropriadamente por meio de intervalo de estimativa e intervalo de confiança amplo ao redor da estimativa de efeito de tratamento, descrever-se-á o baixo poder. O relato do intervalo de confiança representa uma contribuição para evitar problemas no sentido de que “a falta de evidência não é evidência da falta”, por meio de resultados forjados baseados em conclusões simplificadas, com  $p > 0,05$ ;
3. *Os estudos com baixo poder devem ser publicados, tornando-se avaliáveis para a metanálise* – É necessário determinar e publicar o viés, permitindo à comunidade científica a possibilidade de executar grandes estudos para corrigi-los. Estudos não totalmente publicados são chamados de “não-científicos” e “não-éticos” na literatura científica<sup>9,10</sup>.

## MODIFICANDO O TAMANHO DA AMOSTRA

Com adição de participantes avaliados e flexibilidade na pesquisa, os investigadores podem considerar a estratégia de modificação do tamanho da amostra. Geralmente, as modifi-

cações ocasionam aumento do tamanho da amostra. Qualquer modificação no tamanho da amostra em um estágio do estudo deve depender do planejamento a fim de evitar viés.<sup>1,10</sup>

## O QUE OS LEITORES DEVEM PROCURAR NOS CÁLCULOS DO TAMANHO DA AMOSTRA?

Primordialmente, os leitores devem encontrar a estimativa do tamanho da amostra. Em um estudo, o intervalo de confiança indica apropriadamente o poder. Entretanto, o cálculo do tamanho da amostra ainda providencia informações importantes, como<sup>1,10</sup>:

- especificar o desfecho primário com os resultados de segurança e reivindicar o efeito de um resultado não planejado como desfecho primário;
- demonstrar problemas no recrutamento e verificar se o estudo foi interrompido precocemente em razão de resultados estatisticamente significativos de forma adequada;
- considerar erros sistemáticos (viés) realizados pelos investigadores, como randomização pobre, cancelamento de alocação inadequada, cegamento inadequado ou perda de seguimento;
- encontrar, na descrição do cálculo de tamanho da amostra, a descrição do erro tipo I (alfa), o poder (beta), a taxa de evento no grupo controle e o efeito do tratamento (ou, analogicamente, a taxa de evento no grupo tratado). O relato de que “calculamos o tamanho de amostra necessário de X em cada grupo com  $\alpha = 0,05$  e poder = 0,90” é quase insignificante, porque negligencia a estimativa do tamanho da amostra e a taxa de evento no grupo controle.

## CONCLUSÃO

Os estudos não podem ser considerados antiéticos com base isoladamente nos processos de cálculos do tamanho da amostra subjetivos e imprecisos. No entanto, o planejamento para um poder adequado e a execução de grandes estudos multicêntricos devem sempre ser encorajados.

Considerando que o cenário da pesquisa clínica nem sempre permite insistir isoladamente em estudos de larga escala, muitas questões não são respondidas e podem se distanciar de respostas relevantes. A ocasional substituição da ênfase na fixação do tamanho da amostra pelo foco na qualidade metodológica permitiria a execução de estudos com menos vieses. Nesse contexto, estudos sem vieses com resultados precisos podem responder a perguntas ainda relevantes no cenário obscuro da falta de evidência.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. Schulz KF, Grimes DA. Ample size calculations in randomised trials: mandatory and mystical. *Lancet* 2005;365:1348-53.
2. Freiman JA, Chalmers TC, Smith Jr. H, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial: survey of 71 "negative" trials. *N Engl J Med* 1978;299:690-4.
3. Sackett DL, Cook DJ. Can we learn anything from small trials? *Ann NY Acad Sci* 1993;703:25-31.
4. Antes G, Chalmers I. Under-reporting of clinical trials is unethical. *Lancet* 2003;361:978-9.
5. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet* 2002;359:515-9.
6. Schulz KF, Grimes DA. Unequal group sizes in randomised trials: guarding against guessing. *Lancet* 2002;359:966-70.
7. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
8. Dahmen G, Rochon J, König IR, Ziegler A. Sample size calculations for controlled clinical trials using generalized estimating equations (GEE). *Methods Inf Med* 2004;43(5):451-6.
9. Cornelissen VA, Fagard RH. Effect of resistance training on resting blood pressure: a meta-analysis of randomized controlled trials. *J Hypertens* 2005;23(2):251-9.
10. Schulz KF. Subverting randomization in controlled trials. *JAMA* 1995;274:1456-8.