

# Medicina baseada em evidências e hipertensão arterial

## Medicine based on evidences and hypertension

Editor: Álvaro Avezum

## Análise de subgrupos em estudos clínicos randomizados

### Subgroup analysis in randomized clinical trials

Hélio Penna Guimarães<sup>1</sup>, Lígia Nasi Laranjeira<sup>2</sup>, Álvaro Avezum<sup>3</sup>, Leopoldo Soares Piegas<sup>4</sup>

### INTRODUÇÃO

As análises de subgrupos têm um apelo especial: parecem lógicas, intuitivas, promissoras e até mesmo divertidas para investigadores e leitores, entretanto sua multiplicidade e ingenuidade combinadas podem propiciar graves erros de interpretação, condução e relato do estudo<sup>1</sup>.

Quanto maior o número de subgrupos analisados em uma amostra, maior a probabilidade de resultados falso-positivos; soma-se também a essa distocia a freqüente tendência ao relato, após diversas análises, apenas dos resultados significantes que potencialmente distorcem a literatura médica.

A despeito dessas limitações, a prática de análises de subgrupos em estudos clínicos é freqüente. Pocock *et al.*<sup>2</sup>, em análise aleatória de 45 ensaios, descreveram que pelo menos um subgrupo de distinta categoria foi avaliado em 51% destes, sem qualquer prévia predeterminação de análise em seus objetivos. A justificativa mais utilizada para tais análises é que os subgrupos podem aproximar o cenário da prática clínica real, mas será que forneceriam mais informações efetivas ou apenas maior confusão de dados; e em caso de resultado positivo, será que efetivamente há importante diferença no efeito do tratamento ou trata-se de mera ocorrência ao acaso (em se considerando a ausência de prévia determinação da amostragem e conseqüente poder estatístico para tal diferença)<sup>3</sup>.

Essa multiplicidade, se não previamente determinada, pode encorajar o equívoco na interpretação dos resultados, e os efeitos de tratamento em subgrupos relatados podem ser ilusórios.

### ANÁLISE DE SUBGRUPOS: COMO INTERPRETÁ-LA?

A análise indiscriminada de subgrupos oferta multiplicidade de conceitos; este é um freqüente problema de uso excessivo na literatura médica, a despeito de informações e recomendações para sua não-utilização ou uso racional<sup>4,5</sup>.

Com freqüência, os investigadores definem subgrupos dos participantes por suas características basais e realizam análises para avaliar se os efeitos do tratamento diferem nesses subgrupos, em geral utilizando testes estatísticos por vezes equivocados. Questiona-se se efetivamente em cada tratamento ou conduta avaliada haveria diferença entre:

- tipos de pacientes (exs.: idade e sexo);
- tratamentos ou intervenção (doses dentro da mesma classe de medicação);
- possível influência nos desfechos.

Neste cenário, argumentos pró e contra-análises, se aplicáveis, poderiam ser: análise de subgrupos excessiva e não previamente especificada, dita "pescaria" de dados, e exercícios de *data-dredging* podem promover maior confusão de resultados que conclusões relevantes<sup>3,6</sup>. No entanto, o encontro de diferenças poderia alarmar para o risco de não se identificar importantes diferenças no efeito, concluindo-se resultados pela "média" de efeitos na população geral, desconsiderando a heterogeneidade desta.

Desta forma parecem objetivas as seguintes questões: quando acreditar na diferença entre subgrupos? Como medir seu efeito e escolher a correta medida?

1 Médico-assistente da Divisão de Pesquisa do Instituto Dante Pazzanese de Cardiologia.

2 Coordenadora de Estudos Clínicos da Divisão de Pesquisa do Instituto Dante Pazzanese de Cardiologia.

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

4 Diretor técnico do Instituto Dante Pazzanese de Cardiologia.

Correspondência para: Hélio Penna Guimarães. Unidade de Pesquisa Clínica e Epidemiológica (Upec) da Divisão de Pesquisa do Instituto Dante Pazzanese de Cardiologia. Av. Dante Pazzanese, 500 – 1º andar – 04012-909 – São Paulo, SP. Fone/Fax: (11) 5085-6240. E-mail: heliopg@dantepazzanese.org.br

Para tal solução, algumas relevantes perguntas devem ser respondidas pelo estudo:

- Quando as diferenças entre subgrupos seriam reais?

É fundamental determinar se a hipótese precedeu a análise do subgrupo ou não. A credibilidade de qualquer interação deve ser sempre reavaliada se dados foram obtidos após a análise *post hoc* ou não previamente definida<sup>2,7</sup>.

Se, no entanto, a hipótese ou suposta conclusão gerada for sugerida de forma não equivocada por dados diferentes, talvez possa ser considerada para teste em um cenário adequado de um ensaio clínico randomizado.

- O efeito foi testado em subgrupo de pequeno número?

Comumente as hipóteses *post hoc* baseadas na análise do subgrupo incorrem em erro por partir da exploração de dados globais em múltiplas hipóteses e condições e, por conseguinte, menor tamanho de amostra e mais chance do acaso<sup>7</sup>. Nesses casos, ainda que pré-especificados, os subgrupos têm a força da inferência de várias outras hipóteses diminuída de um real achado (pela influência das múltiplas co-intervenções da amostra global). Desta forma é comum que um teste-padrão consistente com a diferença observada entre os vários subgrupos deva-se apenas ao erro de amostragem, propiciando que pequenos tamanhos de amostra demonstrem diferenças maiores em efeito aparente por puro *play of chance*<sup>1,3,7</sup>.

- O efeito mensurado estatisticamente é significativo?

Se o tamanho da amostra for menor, como descrito anteriormente, o efeito não alcançará o significado estatístico desejado (refletido no intervalo da confiança). Desta forma é comum que o investigador busque técnicas estatísticas para explicar diferenças aparentes do subgrupo e caberá ao leitor avaliar se o teste estatístico está apto a demonstrar a real diferença entre os subgrupos. Desta forma, se há grande número de hipóteses ou subgrupos para análise, o risco de valores baixos de *p* é elevado e a chance do acaso para comprovação de distintas hipóteses sempre será maior<sup>1,3,8</sup>.

- Há interação consistente entre o resultado encontrado e estudos prévios?

Um achado absolutamente novo e distinto a prévias tendências ou resultado de outros estudos pode denotar a inconsistência do resultado. Uma revisão sistemática rigorosa da literatura poderia demonstrar credibilidade do achado<sup>9,10</sup>.

- Há pelo menos evidência indireta que suporte a hipótese?

Se houver pelo menos evidência indireta para a hipótese (estudos animais ou situações análogas em humanos), pode ser que se esteja diante de uma hipótese plausível a ser testada em um novo estudo de forma metodológica adequada. Essa evidência para a consistência pode-se basear até mesmo na compreensão dos mecanismos biológicos da doença com estudos de populações diferentes (estudos animais), observações das interações para intervenções similares e resultados de outros estudos relacionados (ainda que intermediários)<sup>9-11</sup>.

- Como aplicar testes estatísticos em subgrupos?

Para análise apropriada, deve-se utilizar testes estatísticos de interação, que envolvam em sua avaliação a hipótese de o efeito do tratamento depender do subgrupo do participante. Isso testa não somente a hipótese, mas também aborda de forma apropriada o problema de multiplicidade. No entanto, há investigadores que questionam testes de interação baseados na falta de poder, entretanto tais testes fornecem o cuidado apropriado, reconhecem a informação limitada disponível nos subgrupos e emergem como o método estatístico mais eficaz para detectar os achados impróprios do subgrupo, ao ainda ter a habilidade de detectar efeitos interativos, se presentes<sup>12,13</sup>.

Os leitores devem ser cautelosos ante as experimentações de relatos com muitas análises de subgrupo, a menos que os investigadores forneçam razões válidas; o ensaio clínico também terá maior crédito se houver indicação clara das análises executadas, sejam positivas ou não em seus achados. Deve-se também denominar quais análises não foram pré-especificadas como hipótese; estas não deverão constar nas conclusões do estudo<sup>13</sup>.

## CONCLUSÃO

Os critérios sugeridos nesta breve revisão para análise de subgrupos são úteis para decidir a plausibilidade de resultado de um subgrupo. Desta forma, para se considerar as hipóteses ou interações geradas como plausíveis, sugere-se confrontar evidências indiretas com diretas prévias, definitivamente comprová-las em novo estudo ou em metanálise. Se os benefícios potenciais do resultado são grandes e os riscos e custos pequenos para aderência são aceitáveis, pode ser conveniente considerar o resultado a despeito do baixo padrão da evidência oferecido. Se, no entanto, parece claro que o mecanismo biológico ou exigência dos patrocinadores, público e indústria claramente poderiam influenciar o achado, deve-se considerar apenas a análise confinada a subgrupos pré-especificados e em número limitado.

## REFERÊNCIAS

1. Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim. Analyses. *Lancet* 2005; 365:1657-61.
2. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials: a survey of three medical journals. *N Engl J Med* 1987;317:426-32.
3. Oxman A, Guyatt G. When to believe a subgroup analysis. In: Guyatt G. *User's guides to medical literature. A manual of evidence based clinical practice*. New York: Jama Press, 2003.
4. Feinstein AR. *Clinical epidemiology: the architecture of clinical research*. Philadelphia: WB Saunders, 1985; 306-307, 516-7.
5. Altman DG. Within trial variation – A false trail? *J Clin Epidemiol* 1998;51:301-3.
6. Feinstein AR. The problem of cogent subgroups: a clinicostatistical tragedy. *J Clin Epidemiol* 1998;51:297-9.
7. Furberg CD, Morgan TM. Lessons from overviews of cardiovascular trials. *Stat Med* 1987;6:295-306.
8. Horwitz RI. Complexity and contradiction in clinical trial research. *Am J Med* 1987;82:498-510.
9. Buyse ME. Analysis of clinical trial outcomes: some comments on subgroup analyses. *Control Clin Trials* 1989;10(suppl 4):187S-194S.
10. Bulpitt CJ. Subgroup analysis. *Lancet* 1988;2:31-4.
11. Byar DP. Assessing apparent treatment – Covariate interactions in randomized clinical trials. *Stat Med* 1985;4:255-63.
12. Shuster J, van Eys J. Interaction between prognostic factors and treatment. *Control Clin Trials* 1983;4:209-14.
13. Schneider B. Analysis of clinical trial outcomes: alternative approaches to subgroup analysis. *Control Clin Trials* 1989;10(suppl 4):176S-186S.