

Requerimentos mínimos para o planejamento e análise de estudos clínicos de não inferioridade*

Minimum requirements for planning and analyzing non-inferiority clinical trials

Andre D. Sasse¹, Cílicia Y. Wada², Joanlise M. Leon Andrade³, Jose V. Bermudez⁴, Paulo A. Lotufo⁵, Valdair F. Pinto⁶

*Recebido do Grupo Multidisciplinar para Discussão de Estudos de Não Inferioridade. São Paulo, SP.

RESUMO

JUSTIFICATIVA E OBJETIVOS: Estudos de não inferioridade, de uso crescente em pesquisa clínica, são metodologicamente distintos dos estudos tradicionais de superioridade e são particularmente úteis quando existem impedimentos éticos para estudos placebo controlados. O objetivo deste estudo foi rever os critérios mínimos a serem observados no planejamento, condução e análise dos estudos de não inferioridade.

CONTEÚDO: Os seguintes determinantes críticos dos estudos de não inferioridade são apresentados e discutidos: margem de não inferioridade, sensibilidade de ensaio, tamanho da amostra, estratégias de análise, algoritmo de inferência e condições que suportam conclusão de superioridade.

CONCLUSÃO: Os estudos de não inferioridade possuem um conjunto particular de vieses a serem controlados e o uso correto desta metodologia é orientado por algumas diretrizes regulatórias internacionais. A observância de certo número de requerimentos mínimos é condição necessária para assegurar a validade das conclusões dos estudos de não inferioridade.

Descritores: Delineamentos experimentais, Ensaios clínicos, Estudos de não inferioridade.

SUMMARY

BACKGROUND AND OBJECTIVES: Non-inferiority trials with increasing utilization in clinical research have different properties in relation to traditional superiority trials. They are particularly useful in situations where placebo controlled studies are considered unethical. This study is a brief review of basic criteria to be observed for planning, executing and analyzing non-inferiority studies.

CONTENTS: The following critical issues of non-inferiority studies are discussed: non-inferiority margin, assay sensitivity, sample size determination, strategies for analysis, rules of inference and conditions for switching to claim of superiority.

CONCLUSION: Non-inferiority designs are subject to a particular set of bias and some international guidelines are available. Clinical investigators should be aware of basic requirements to ensure validity of conclusions from non-inferiority studies.

Keywords: Clinical trials, Non-inferiority trials, Study design.

INTRODUÇÃO

Estudos clínicos de não inferioridade são estudos delineados para avaliar se um novo tratamento não é menos eficaz do que um tratamento padrão, por mais de uma margem de tolerância. Esta margem deve ser fixada previamente e é denominada “margem de não inferioridade”. Tais estudos não devem ser confundidos com os *estudos de equivalência*, que visam avaliar a ausência de diferenças importantes entre tratamentos, também dentro de uma margem pré-definida. Por outro lado, os *estudos de superioridade* são aqueles com delineamento mais tradicional e têm por objetivo avaliar se um novo tratamento é mais eficaz que um controle (placebo ou substância ativa).

As bases metodológicas e fundamentos estatísticos para o planejamento e análise dos estudos clínicos de equivalência foram estabelecidos no final da década de 1970¹. Refinamentos posteriores levaram à formulação dos estudos clínicos de não inferioridade, que se apresentaram como alternativas válidas para a comparação de novos tratamentos com controles ativos, especialmente em circunstâncias em que o uso de placebo seria inviável. Os fundamentos elaborados definiram diferenças claras para os clássicos estudos clínicos de superioridade. Com frequência constituem desafios para pesquisadores clínicos, bioestatísticos, editores de revistas médicas e agências regulatórias. Algumas diretrizes internacionais específicas para estudos de não inferioridade, já foram

1. Médico do Centro de Evidências em Oncologia da Universidade Estadual de Campinas (UNICAMP) e Radium Instituto de Oncologia. Campinas, SP, Brasil

2. Livre-Docente do Instituto de Matemática e Estatística da Universidade Estadual Paulista. Diretora da Statpharm Consultoria. Campinas, SP, Brasil

3. Professor-Adjunto do Departamento de Estatística da Universidade de Brasília. Brasília, DF, Brasil

4. Médico-Pesquisador do Instituto de Infectologia Emilio Ribas da Faculdade de Medicina de São Paulo (FMUSP). São Paulo, SP, Brasil

5. Livre Docente de Clínica Médica e Epidemiologia Cardiovascular e Superintendente do Hospital Universitário da Universidade de São Paulo. São Paulo, SP, Brasil

6. Consultor Independente em Medicina Farmacêutica, Ex-Professor de Bioestatística da Faculdade de Medicina da Universidade do Triângulo Mineiro, Uberaba, MG, Brasil

Apresentado em 21 de junho de 2010

Aceito para publicação em 14 de setembro de 2010

Endereço para correspondência:

Valdair F. Pinto

Alameda dos Aicas, 229

04086-000 São Paulo SP.

E-mail: valdairpinto@hotmail.com

© Sociedade Brasileira de Clínica Médica

publicadas²⁻⁶. No entanto, ainda não estão disponíveis orientações básicas para a condução deste tipo de estudo clínico. Este estudo foi resultado de discussões de um grupo multidisciplinar com interesse e experiência em pesquisa clínica e teve como objetivo indicar critérios mínimos que devem ser observados no planejamento e análise de estudos clínicos de não inferioridade.

REQUERIMENTOS MÍNIMOS E RECOMENDAÇÕES

Escolha de margem de não inferioridade

A margem de não inferioridade deve ser estabelecida antes do início do estudo, especificada no protocolo e justificada clínica e estatisticamente. Como princípio, a margem deve ser menor que uma diferença clinicamente aceitável e também, menor que o efeito estimado integral do controle em relação ao placebo⁵.

Há várias maneiras para se escolher a margem de não inferioridade. O mais utilizado é o chamado *método de margem fixa* que se baseia na estimativa do efeito do controle em relação ao placebo (de estudos anteriores), seguido da decisão de qual proporção deste efeito se deseja preservar no tratamento em estudo para que ele seja declarado não inferior⁷. A sequência de eventos para esta determinação pode ser assim resumida:

- Assegurar que exista mais de um estudo, com delineamentos similares, comparando o controle e placebo;
- Avaliar os intervalos de confiança de 95% (IC-95) de cada estudo e tomar o menor limite inferior dos IC-95;
- Utilizar uma fração deste valor (uma estimativa conservadora do efeito do controle) para fixar a margem de não inferioridade (por exemplo, 50% do limite inferior do IC-95)

Outro enfoque, mais complexo, chamado *método sintético*, combina através de metanálise as estimativas do efeito do controle encontradas em estudos prévios com o efeito do controle no estudo atual. Desta maneira é também possível comparar o efeito do novo tratamento com uma estimativa projetada do efeito placebo (placebo putativo), mesmo na ausência de um grupo placebo real^{8,9}.

Qualquer que seja o método, o objetivo é garantir que o efeito do tratamento em estudo, sendo não inferior ao controle, seja necessariamente superior ao placebo.

Sensibilidade de ensaio

Sensibilidade de ensaio (SE) é a propriedade de um estudo clínico de detectar diferenças entre tratamentos quando elas de fato existem³. A ausência de SE pode introduzir um viés de não diferença (p. ex.: aumento na taxa de falsos negativos) nas comparações, causado pelo aumento no erro tipo 2. Esse viés, que em estudos de superioridade pode resultar no fracasso em se demonstrar eficácia, em estudos de não inferioridade pode levar tratamentos ineficazes a serem declarados não inferiores. Portanto, a ausência de SE tem repercussões ainda mais sérias, do ponto de vista clínico, num estudo de não inferioridade. Vários fatores podem reduzir a SE tais como: não adesão ao tratamento, critérios de inclusão permissivos, maior variabilidade de medidas, etc. É importante notar que a SE é um conceito mais amplo que o poder estatístico do estudo, pois inclui a qualidade de execução além de outras circunstâncias metodológicas.

Idealmente, a melhor maneira de se controlar a SE é pela inclusão de um terceiro grupo com placebo, permitindo a validação inter-

na da sensibilidade. Entretanto, como o uso de placebo muitas vezes é inviável, a SE pode ser deduzida ou presumida através de validação externa, utilizando informações de controles históricos. Assim, controlar a SE significa assegurar que no estudo em questão o efeito do grupo controle seja superior ao placebo, caso este tivesse sido incluído.

A existência de SE pode ser deduzida a partir de três condições: evidência histórica de sensibilidade de efeito, pressuposto de constância e qualidade de execução do estudo⁸:

- Evidência histórica de sensibilidade de efeito de tratamento, conhecida pela sigla em inglês HESDE (*Historical evidence of sensitivity to drug effects*), significa que estudos clínicos adequadamente planejados e conduzidos no passado mostram que o tratamento, aqui considerado como controle, é regular e consistentemente superior ao placebo. Esta é uma conclusão que depende da patologia, do tratamento e das condições experimentais. HESDE é uma propriedade que não é facilmente assumida para, por exemplo, antidepressivos, analgésicos e tratamento para insônia. Ao contrário, é mais facilmente admitida para anticoagulantes, antibióticos e antineoplásicos, por exemplo;
- Pressuposto de constância é a condição em que a eficácia do controle em relação a placebo é mantida para estudos similares. A evidência de constância permite assumir que a mesma superioridade do controle em relação a placebo deve ser mantida no estudo atual. O objetivo é garantir que o controle não tenha eficácia volátil, produzindo efeitos variáveis;
- Qualidade de execução do estudo é avaliada através da identificação para cada estudo, de uma lista de fatores que potencialmente possam reduzir diferenças entre tratamentos e/ou aumentar a variabilidade, causando viés de não diferença. Um gerenciamento destes fatores deve ser indicado no protocolo.

Se estas condições não estão presentes, deve-se considerar a inclusão de um grupo placebo. Se isto não é possível, a realização do estudo de não inferioridade pode não estar indicada.

Tamanho da amostra e inferência

O cálculo do tamanho da amostra para estudos de não inferioridade segue o mesmo princípio geral dos estudos clássicos de superioridade, com a inclusão da margem de não inferioridade M . Para o cálculo, sob a condição da hipótese alternativa, podem-se assumir como iguais os verdadeiros valores para os tratamentos e assim a diferença clinicamente relevante, usada no caso dos estudos de superioridade, é substituída pela margem de não inferioridade¹⁰. O tamanho da amostra é muito sensível às variações da margem porque o número de pacientes necessário em estudos de não inferioridade é inversamente proporcional ao quadrado do valor da margem. Portanto reduzindo-se a margem pela metade, quadruplica-se o número necessário de pacientes.

A determinação adequada do tamanho da amostra tem repercussões econômicas e éticas. Um número insuficiente de pacientes para que se atinja o poder do teste pré-especificado desperdiça recursos e expõe os pacientes desnecessariamente à situação de pesquisa por ser finalizado com resultados inconclusivos. O tamanho da amostra deve ser justificado no protocolo e prever ajustes para perdas de seguimento e comparações múltiplas, quando necessárias.

Os estudos de não inferioridade podem ser analisados pelo tradi-

cional teste de hipóteses. No entanto, o método de escolha para a interpretação dos resultados deve ser por intervalos de confiança. No primeiro caso a hipótese nula é submetida a um teste unilateral com alfa de 0,025. No segundo caso, mais usual e recomendado pelas autoridades regulatórias, são construídos intervalos de 95% de confiança, bilaterais (alfa de 0,05). O tratamento em estudo é declarado não inferior se o limite inferior do IC-95 da diferença entre tratamento e controle não incluir o valor da margem de não inferioridade especificada¹¹.

Conjunto de dados para análise

Idealmente todos os pacientes incluídos em um estudo deveriam completá-lo cumprindo todos os critérios estabelecidos no protocolo inicial. Na prática, esta situação quase nunca acontece, por motivos diversos. Comumente ocorrem perdas, faltas de adesão aos tratamentos e outros desvios do protocolo. Para a análise dos resultados são definidas duas populações básicas: a primeira é definida pela “*intenção-de-tratamento*” (ITT) que inclui todos os pacientes independentemente de violações de protocolo e perdas de seguimento; a segunda, referida como “*por-protocolo*” (PP), subconjunto da primeira, inclui somente os pacientes que completaram o tratamento de acordo com os critérios do protocolo, excluindo todos os violadores¹².

Quando há violação significativa dos critérios estabelecidos no protocolo (muitas perdas, troca ou pouca adesão ao tratamento, e/ou outros desvios) a análise por ITT pode incluir um viés de não diferença resultando em menor efeito de tratamento e maior risco de se declarar falsamente não inferioridade. Já a análise PP, nessas situações adversas, pode incluir um viés de diferença, resultando maior efeito dos tratamentos e maior risco de se declarar falsamente superioridade. Portanto, as análises ITT são conservadoras para superioridade enquanto que as PP são conservadoras para não inferioridade^{13,14}.

Pelas razões acima mencionadas, o conjunto ITT é considerado padrão para estudos de superioridade, enquanto que o conjunto PP é preferido para estudos de não inferioridade. Entretanto, para diminuir o risco de falsas inferências as autoridades regulatórias internacionais (FDA e EMEA) recomendam que os estudos de não inferioridade sejam analisados pelos dois conjuntos de dados, sendo necessário que as conclusões resultantes sejam similares e consistentes.

Conclusão de superioridade

A definição do tipo de estudo, se superioridade ou não inferioridade, deve ser feita obviamente antes do início estudo, pois dela dependem vários elementos do planejamento, como escolha do comparador, margem de não inferioridade e tamanho da amostra. Não obstante, os resultados de um estudo planejado como não inferioridade podem sugerir superioridade e é lícito reportá-lo como superioridade, sem correções ou ajuste no valor de alfa, obedecendo às seguintes condições^{4, 15}:

- O intervalo de confiança IC-95 da diferença entre tratamento e controle situa-se inteiramente acima da margem e também acima de zero;
- A magnitude do efeito representa relevância clínica, pois o planejamento como não inferioridade inclui um número maior de pacientes que potencialmente pode detectar diferenças clínica-

mente irrelevantes (excesso de potência);

- A conclusão de superioridade foi alcançada com uma análise seguindo o princípio de intenção-de-tratamento (ITT);
- A superioridade não é perdida em variáveis secundárias, como efeitos adversos.

A situação inversa, isto é concluir não inferioridade a partir de um estudo que falhou em mostrar superioridade, é muito mais difícil e encontra um grande número de restrições. Para fins práticos é quase impossível, pois no planejamento esta alternativa deveria ter sido prevista com definição de margem, determinação de sensibilidade, HESDE, pressuposto de constância e tamanho da amostra, o que em geral quase nunca acontece.

CONCLUSÃO

O uso correto da metodologia de não inferioridade requer a observância de um conjunto de critérios básicos e controles de vieses específicos. Finalmente, deve-se mencionar que os estudos de não inferioridade possuem limitações inerentes e específicas como impossibilidade de validação interna de sensibilidade, justificativas algumas vezes arbitrarias para margens e falta de proteção adequada contra viés de não diferença. Para elaborar e/ou revisar relatórios de estudos de não inferioridade alguns “checklists” estão disponíveis. Dois destes são especialmente úteis, um elaborado pelo grupo CONSORT⁶ e outro pelo *Canadian Partnership Against Cancer* da Universidade de McMaster¹⁶. Estes instrumentos simples oferecem uma forma clara e estruturada de examinar os diferentes elementos dos estudos de não inferioridade.

Apêndice

Algoritmo para análise estatística

Seja **T** o valor da variável de eficácia do novo tratamento e **C** a variável de eficácia do tratamento controle. A comparação entre estas medidas pode ser expressa como uma diferença entre médias ou proporções ou, em eventos de resposta binária, como uma relação entre riscos ou razão de chances (*odds ratio*) e em estudos de sobrevida pode ser uma relação entre taxas de riscos (*hazard ratio*).

Defina θ como a diferença ou a razão entre as medidas de **T** e **C** e a margem de não inferioridade como **M**.

As hipóteses nula (H_0) e alternativa (H_a) são:

$H_0: \theta \leq M$ (T é inferior a C)

$H_a: \theta > M$ (T é não inferior a C)

onde $M < 0$

Implementa-se o teste de hipóteses unilateral com nível de significância α ou calcula-se o Intervalo de Confiança bilateral CI100(1-2 α).

Se SE_θ o erro-padrão de θ e z_α o percentil 100. α % da distribuição normal padrão, a não inferioridade é declarada se:

$$z > z_\alpha \text{ onde } z = (\theta - M)/SE_\theta$$

ou

se o limite inferior de CI100(1-2 α) > M

Os dados devem suportar o pressuposto de que a estatística z tem assintoticamente distribuição normal na condição da hipótese nula.

REFERÊNCIAS

1. Dunnett CW, Gent M. Significance testing to establish equivalence between treatments, with special reference to data in the form of 2X2 tables. *Biometrics* 1977;33(4):593-602.
2. International Conference of Harmonization (ICH). Guideline E9 - Statistical principles for clinical trials. Fevereiro 1998. Disponível em <http://www.ich.org> (acesso em 20/4/2010).
3. International Conference of Harmonization (ICH). Guideline E10 - Choice of control groups and related design and conduct issues in clinical trials. Julho 2000. Disponível em <http://www.ich.org>. (acesso em 20/4/2010).
4. Committee for Proprietary Medicinal Products. **Points to consider** on switching between superiority and non-inferiority. EMEA/CMPM/EWP/482/99. July 2000. Disponível em <http://www.emea.europa.eu>. (acesso em 20/4/2010).
5. Committee for Proprietary Medicinal Products. Guidelines on the choice of the non-inferiority margin. EMEA/CMPM/EWP/2158. July 2005. Disponível em <http://www.emea.europa.eu>. (acesso em 20/4/2010).
6. Piaggio G, Elbourne DR, Altman DG, et al. Reporting of noninferiority and equivalence randomized trials: an extension of CONSORT statement. *JAMA* 2006;295(10):1152-160.
7. Tsong Y, Zhang J, Levenson M. Choice of delta noninferiority margin and dependency of the noninferiority trials. *J Biopharm Stat* 2007;17(2):279-88.
8. D'Agostino RB Sr, Massaro JM, Sullivan SM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Stat Med* 2003;22(2):169-86.
9. Snapinn SM. Alternatives for discounting in the analysis of noninferiority trials. *J Biopharm Stat* 2004;14(2):263-73.
10. Lesaffre E. Superiority, equivalence and non-inferiority trials. *Bull NYU Hosp Jt Dis* 2008;66(2):150-4.
11. Pater C. Equivalence and noninferiority trials - are they viable alternatives for registration of new drugs? *Curr Control Trials Cardiovasc Med* 2004;5(1):8-14.
12. Matilde Sanchez M, Chen X. Choosing the analysis of population in non-inferiority studies: per protocol or intent-to-treat. *Stat Med* 2006;25(7):1169-81.
13. Christensen E. Methodology of superiority vs. equivalence trials and non-inferiority trials. *J Hepatol* 2007;46(5):947-54.
14. Fueglistaler P, Adamina M, Guller U. Non-inferiority trials in surgical oncology. *Ann Surg Oncol* 2007;14(5):1532-9.
15. Zee BC. Planned equivalence or noninferiority trials versus unplanned noninferiority claims: are they equal? *J Clin Oncol* 2006;24(7):1026-8.
16. Canadian Partnership Against Cancer. Capacity Enhancement Program: non-inferiority and equivalence trials checklist. *McMaster University*. 2009. Disponível em: <http://fhs.mcmaster.ca/cep/documents/CEPNon-inferiorityandEquivalenceTrialChecklist.pdf> (acesso em 20/4/2010).