

Estudos clínicos de não-inferioridade: fundamentos e controvérsias

Non-inferiority clinical trials: concepts and issues

Valdair Ferreira Pinto*

Resumo

Estudos de não-inferioridade são modelos experimentais desenvolvidos com o objetivo de determinar se um novo tratamento ou procedimento não é menos eficaz que outro já estabelecido e considerado como controle. São de grande importância no estudo de tratamentos em que o uso de placebo é inviável. Requerem metodologia diferente dos estudos clássicos, chamados de estudos de superioridade, especialmente no planejamento e análise estatística. O presente artigo é uma revisão dos principais elementos diferenciais entre estudos de não-inferioridade e estudos tradicionais. Existe ainda em nosso meio um considerável grau de desconhecimento sobre o uso correto dessa metodologia, o que certamente compromete a validade de alguns estudos clínicos.

Palavras-chave: Ensaio clínico, delineamentos experimentais, não-inferioridade.

Abstract

Non-inferiority trials are experimental models designed to determine whether a new treatment or procedure is not less effective than an established one, which is considered as standard. They are especially important in the assessment of treatments in which the use of placebo is impracticable. They differ substantially from the classical superiority trials and require different approach, especially in the planning and the data analysis. This paper is a review of the key differences between non-inferiority and traditional clinical studies. There is a considerable amount of misunderstanding on the correct use of this experimental design, which certainly compromises the credibility of some clinical assessments.

Keywords: Clinical trial, Research Design, statistics & numerical data.

Introdução

Os ensaios clínicos comparativos randomizados são considerados atualmente o melhor delineamento experimental para avaliar questões relacionadas a tratamento e prevenção¹. Classicamente, são definidos como experimentos médicos projetados para determinar qual de duas ou mais intervenções é a mais eficaz mediante a alocação aleatória, isto é, randomizada, dos pacientes aos diferentes grupos de estudo. Em geral, um dos grupos é considerado controle – o que algumas vezes pode ser ausência de tratamento, placebo, ou mais frequentemente, um tratamento de eficácia reconhecida. Recursos estatísticos são disponíveis para validar conclusões e maximizar a chance de identificar

o melhor tratamento. Esses modelos são chamados de estudos de superioridade, cujo objetivo é determinar se um tratamento em investigação é superior ao agente comparativo. Algumas alternativas de alocação não-aleatória de pacientes foram propostas para minimizar a variância entre grupos e aumentar eficiência e sensibilidade do estudo; alguns desses métodos têm sólidos fundamentos estatísticos^{2,3}; no entanto, os procedimentos de randomização por blocos com ou sem estratificação ainda são os modelos mais utilizados. Um frequente erro de lógica cometido na interpretação dos estudos clínicos tradicionais é o de admitir equivalência dos tratamentos quando não foi possível mostrar diferenças significativas, isto é, a não-significância em um teste tradicional não suporta a conclusão de não-diferença

Valdair Pinto Consultoria em Medicina Farmacêutica, São Paulo, SP.

* Consultor em Medicina Farmacêutica; Ex-professor de Bioestatística da Faculdade de Medicina da Universidade do Triângulo Mineiro, Uberaba, MG; Ex-diretor médico internacional da Pfizer Inc., Nova York, EUA.

O presente artigo é parte de um projeto de educação médica em estudos de não-inferioridade que teve apoio parcial e incondicional da Associação da Indústria Farmacêutica de Pesquisa (Interfarma).

Artigo submetido em 19.04.2010, aceito em 23.06.2010.

J Vasc Bras. 2010;9(3):145-151.

(erro tipo II). Em condições de baixo poder estatístico do estudo, importantes diferenças entre tratamentos podem existir, e a hipótese nula não é rejeitada devido a um número insuficiente de pacientes, à exagerada variabilidade ou vícios de delineamento e/ou de condução do estudo. Esse erro de inferência tem sido, com frequência, ressaltado com a expressão “A ausência de evidencia não é uma evidencia de ausência”^{4,5}.

Como é fundamentalmente impossível demonstrar que dois tratamentos são iguais, a partir da década de 1970 surgiram novos procedimentos metodológicos que permitiram o desenvolvimento dos chamados estudos de equivalência – destinados a mostrar a ausência de diferenças importantes entre tratamentos – e dos estudos de não-inferioridade, de crescente utilização, os quais têm o objetivo de mostrar que um novo tratamento não é, dentro de certos critérios, menos eficaz que outro já existente. Os estudos de equivalência tornaram-se essenciais para a aprovação regulatória dos medicamentos genéricos, enquanto os estudos de não-inferioridade são hoje utilizados em situações nas quais comparações com placebo são inviáveis e controles ativos são necessários⁶. As diferenças entre os estudos de superioridade, equivalência e não-inferioridade estão essencialmente na formulação das hipóteses a serem testadas. A Tabela 1 apresenta os algoritmos de análise para esses três tipos de estudos, em que T representa a medida de eficácia do novo tratamento e C, a medida de eficácia do tratamento controle. Rejeitar a hipótese nula significa, para os estudos de superioridade, que T é superior a C; para os estudos de não-inferioridade, que a diferença entre C e T é menor que uma margem M e, para os estudos de equivalência, que a diferença entre C e T não é menor nem maior que uma margem M. Em essência, o termo equivalente significa não-inferior-e-não-superior, e testar para equivalência remete à análise para a região simétrica definida por [+M , -M], conforme indicado na Figura 1.

Os estudos de não-inferioridade têm grande aplicação em Oncologia, na Cardiologia preventiva e na avaliação de anti-infecciosos. As autoridades regulatórias têm

Tabela 1 - Formulação de hipóteses para estudos de superioridade, não-inferioridade e equivalência

Tipo de estudo	Hipótese nula	Hipótese alternativa
Superioridade	$H_0: C - T \geq 0$	$H_a: C - T < 0$
Não-inferioridade	$H_0: C - T \geq M$	$H_a: C - T < M$
Equivalência	$H_0: C - T \geq M$	$H_a: C - T < M$

T e C são medidas de eficácia para o novo tratamento e controle, respectivamente; M é a margem de não-inferioridade/equivalência

ultimamente requerido também estudos de não-inferioridade na avaliação de biossimilares (produtos obtidos por processos biotecnológicos, como proteínas terapêuticas, hemoderivados, anticorpos monoclonais etc.). Um produto que se mostra não-inferior em relação a um tratamento estabelecido quanto a uma variável de eficácia pode, entretanto, apresentar vantagens importantes, como melhor tolerabilidade, conveniência de uso, vantagens galênicas, diferentes vias de metabolização, menos interações, entre outras.

Uma definição apropriada para os estudos de não-inferioridade deve indicar que estes são destinados a estabelecer se um novo tratamento não é menos eficaz que um tratamento padrão, por mais que uma margem de tolerância fixada previamente e denominada margem de não-inferioridade (M)⁷. Nesses estudos, a hipótese nula é a de que o tratamento em investigação é inferior ao controle por uma diferença maior ou igual a M, e a hipótese alternativa é a de que a diferença entre tratamentos é menor que a margem. O método de escolha para a análise dos estudos de não-inferioridade consiste na construção de intervalos de confiança, usualmente de 95% (IC95%). O tratamento é declarado não-inferior se o limite inferior do IC95% da diferença entre tratamento e controle não incluir o valor da margem especificada.

No planejamento, análise e interpretação dos estudos de não-inferioridade, pelo menos cinco fatores devem ser cuidadosamente considerados para garantir a validade do estudo:

1. escolha da margem de não-inferioridade;
2. número de pacientes necessários ao estudo;

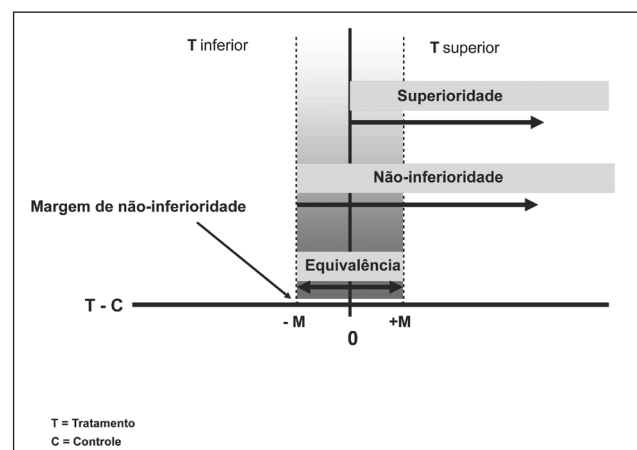


Figura 1 - T: tratamento; C: controle. T é superior a C se o intervalo de confiança da diferença situar-se inteiramente à direita de zero, não-inferior se inteiramente à direita de -M e equivalente se contido na zona de equivalência entre -M e +M

3. controle da sensibilidade do estudo;
4. definição da população de análise;
5. justificativa ética.

Uma breve discussão de cada um desses determinantes críticos foi o objetivo do presente trabalho. Não obstante, o foco deste artigo foram os estudos de não-inferioridade, e muito do que se discute também se aplica aos estudos de equivalência.

Margem de não-inferioridade

Como indicado acima, M quantifica a máxima perda de eficácia clinicamente aceitável para que o tratamento em estudo possa ser declarado não-inferior ao controle. Por conseguinte, não pode exceder a mínima diferença clinicamente relevante que seria utilizada em um estudo de superioridade e , para assegurar a manutenção de alguma eficácia, não pode ser igual ou maior que o efeito integral do tratamento controle. A sua especificação é uma tarefa difícil, porém essencial, sendo também um dos elementos necessários para o cálculo do tamanho da amostra. Valores excessivamente altos para M aumentam a probabilidade de que tratamentos inferiores sejam declarados não-inferiores, enquanto valores pequenos, mais conservadores, exigem amostras maiores, tornando mais dispendiosos os estudos devido a um número maior de pacientes, além de óbvias implicações éticas. O valor de M deve ser estabelecido com base em considerações clínicas e estatísticas e definido *a priori*; pode ser especificado em termos absolutos ou relativos, como diferenças entre médias ou proporções ou logaritmos de razão de chances (odds ratio) ou relações entre taxas de risco (hazard rate).

De uma forma simples, M pode ser fixado como uma porcentagem do efeito estimado do controle no estudo atual, usualmente entre 10 e 20%. A sua definição deve, entretanto, levar em consideração a área terapêutica e a magnitude do efeito do grupo controle; por exemplo, para anti-infecciosos recomenda-se margens mais conservadoras (por exemplo, 10%) quando o efeito esperado é da ordem de 90% e margens mais amplas (por exemplo, 20%) quando o efeito antecipado é inferior a 80%. Deve-se também considerar a existência de outros possíveis benefícios; aceita-se uma margem maior se existem vantagens clínicas como uma redução importante de reações adversas. Situações nas quais o estudo analisa desfechos sólidos, como mortalidade ou morbidade irreversível, a definição de M pode ser muito difícil ou mesmo impossível.

Atualmente, há uma tendência em definir o valor de M predominantemente por meio de considerações estatísticas, deixando o julgamento clínico em uma posição menos decisiva. Assim, utilizando dados históricos de estudos comparando o tratamento controle com placebo, é possível derivar comparações indiretas para assegurar que, mesmo na inexistência de um grupo placebo, o tratamento em estudo seja superior ao placebo. Isso é conhecido como comparação com placebo putativo⁸. Se a magnitude do efeito do controle é designada por C e P é o efeito do placebo baseado em dados históricos, a margem de não-inferioridade M será sempre uma fração de $C-P$, ou seja: $M = x\%$ de $(C-P)$. Dessa maneira, pode-se também controlar quanto do efeito de C se quer preservar em T . Usualmente, aceita-se 50 a 75% como fração a ser preservada do efeito estimado do controle em relação ao placebo. Nesse caso, M seria 25 ou 50% do efeito líquido de C . Um exemplo numérico: suponha que dados históricos indicaram uma eficácia do controle de 80% e de 60% para placebo; o efeito líquido de C será de 20%. Para preservar 50%, M deve ser igual a 10% e para preservar 75%, M deve ser igual 5%. Nesse ponto, considerações clínicas podem ajudar na tomada de decisão e definir um valor final.

A margem de não-inferioridade pode ainda ser decidida pela chamada “50% rule”, endossada pela agência americana Food and Drug Administration (FDA)⁹, a qual prescreve que o valor de M deve ser menor (de preferência 50%) que o limite inferior do IC95% obtido de dados históricos que comparam o tratamento controle com placebo. No exemplo numérico acima, supondo que a diferença de 20% entre as proporções tenha sido obtida de uma amostra de 200 pacientes por grupo, o IC95% é de 11,1- 28,2%. Tomando-se a metade do limite inferior, 11,1%, o valor sugerido para M é de 5,5%. Esse método é considerado conservador, pois dá um duplo desconto no cálculo da margem, diminuindo o poder do estudo em demonstrar não-inferioridade.

Tamanho da amostra

No cálculo do número de pacientes necessários em um estudo clínico, leva-se em consideração a probabilidade alfa (erro tipo I ou falso-positivo) e a probabilidade beta (erro tipo II ou falso-negativo). No contexto dos estudos de não-inferioridade, esses erros têm, de certa forma, interpretação reversa. Alfa quantifica o risco de se declarar falsamente não-inferioridade e pacientes futuros poderão usar um medicamento inferior; beta quantifica o risco de se concluir falsamente pela inferioridade, significando que pacientes futuros podem não se beneficiar do novo medicamento.

Além disso, o tamanho da amostra dependerá também do valor estipulado para a margem de não-inferioridade e da variabilidade dos dados¹⁰.

De uma maneira genérica e simples, a equação para o cálculo do número de pacientes em estudo clínico comparativo por grupos paralelos tem a seguinte forma (Equação 1):

$$n = 2 \times z^2 \frac{s^2}{d^2} \quad \text{Equação 1}$$

onde:

n é o número de pacientes em cada grupo de tratamento;
 z é uma constante que depende das escolhas para alfa e beta, e seus valores são obtidos em tabelas de probabilidades cumulativas da distribuição normal;
 s^2 é uma medida da variabilidade dos dados;
 d é a mínima diferença clinicamente relevante; no caso clássico de alfa = 0,05 e beta = 0,2, o valor de z^2 será igual a 7,9.

Assim, a equação se reduz a (Equação 2):

$$n = 16 \frac{s^2}{d^2} \quad \text{Equação 2}$$

Isso corresponde à elegante sugestão mnemônica de Lehr: "sixteen s-squared over d-squared"¹¹.

Para os estudos de não-inferioridade, é usual estipular alfa = 0,025 em teste unilateral e beta = 0,1 (neste caso, a constante z^2 será igual a 10,5), e d é substituído pela margem de não-inferioridade M (Equação 3).

$$n = 21 \frac{s^2}{M^2} \quad \text{Equação 3}$$

Os métodos para estimar a variância s^2 dependem da natureza da variável em estudo (porcentagens, médias ou taxas de risco), cuja discussão está fora do escopo deste artigo.

Da observação da equação acima, conclui-se que n cresce rapidamente quando M diminui; por exemplo, em uma situação em que a eficácia é estimada em 70%, o número de pacientes necessários em cada grupo será de 83 para $M = 20\%$, de 147 para $M = 15\%$, de 330 para $M = 10\%$ e 1.319 para $M = 5\%$. Quando M tende a zero, n tende a infinito, o que indica ser impossível provar que dois tratamentos são iguais. Mantidas as mesmas propriedades estatísticas e sendo M sempre menor que d (a diferença clinicamente relevante dos estudos de superioridade), o número de pacientes necessários aos estudos de não-inferioridade será sempre maior que o correspondente número para os estudos clássicos de superioridade. Para M igual à metade de d , o número será quatro vezes superior. No caso de estudos de equivalência, o número é ainda ligeiramente maior,

pois são testadas duas hipóteses para a região simétrica $[+M$ e $-M]$, isto é, não-inferioridade e não-superioridade.

O cálculo apropriado do tamanho da amostra é importante não somente para a validade da análise e interpretação como também para o planejamento de recursos; o orçamento de um estudo depende em grande medida do número de pacientes a serem recrutados. Para delineamentos experimentais mais complexos e estratégias de análise mais sofisticadas, os métodos acima descritos são insuficientes e, muitas vezes, simulações podem ser necessárias.

Sensibilidade de ensaio

Sensibilidade de ensaio é a propriedade de um ensaio clínico ser capaz de discriminar um tratamento eficaz de placebo ou, de uma maneira mais geral, é a propriedade de detectar diferenças entre tratamentos quando elas de fato existem¹². A sensibilidade de ensaio depende, entre outros fatores, da magnitude do efeito do tratamento e da qualidade de execução do ensaio, além de fatores como aderência, critérios de seleção de pacientes e do excesso de variabilidade dos dados. É, portanto, um conceito mais abrangente que o poder estatístico do estudo.

Idealmente, um estudo de não-inferioridade deveria incluir um grupo placebo para certificar que, neste estudo em particular, o grupo controle foi eficaz. Como essa situação é quase sempre inviável, uma comparação de não-inferioridade que indica uma diferença de eficácia entre o grupo controle e o tratamento em estudo menor que a margem de não-inferioridade pode significar que os dois tratamentos foram eficazes ou que ambos foram ineficazes. Um estudo de não-inferioridade sem um grupo placebo somente poderá ser validado se for possível assumir que no estudo atual o grupo controle foi eficaz⁶. O pressuposto de eficácia do grupo controle deve ser assumido ou deduzido de forma transitiva, com base em dados históricos. A evidência histórica de regular e consistente eficácia é conhecida pela sigla inglesa HESDE (*historical evidence of sensitivity to drug effects*), que depende principalmente da patologia em estudo e da área terapêutica. A HESDE não pode ser assumida facilmente para drogas que não se mostram regularmente superiores a placebo, tais como antidepressivos, produtos para demência, refluxo gastroesofágico, rinite alérgica etc. Ao contrário, a HESDE pode ser, por exemplo, mais facilmente admitida para anticoagulantes em trombose venosa profunda, antibióticos em infecções urinárias ou beta-agonistas em broncoespasmo. Além da evidência histórica, deve também existir o pressuposto de constância,

garantindo que a evidência de eficácia não variou do passado até o presente¹³.

Métodos mais complexos, chamados de sintéticos¹⁴, procuram combinar erros de duas fontes, dados históricos e estudo atual, construindo uma espécie de erro tipo I “meta-analítico”, o que permite comparar o novo tratamento com o efeito projetado de placebo (placebo putativo). Para não comprometer a validade das conclusões apresentadas, a análise da sensibilidade deve ser sempre conduzida e monitorada a partir de informações externas ao estudo.

Outro fator importante para garantir a superioridade em relação a placebo é a escolha do comparador. Se um tratamento que se demonstrou não-inferior a um padrão reconhecido é posteriormente utilizado como controle em estudos futuros, corre-se o risco de degradar a eficácia e terminar com tratamentos não-superiores a placebo. Tratamentos ligeiramente inferiores, porém dentro da margem, tornando-se controles para uma nova geração de estudos de não-inferioridade, resultam em progressiva perda de confiabilidade. Esse fenômeno da degeneração do comparador é denominado de *biocreep*¹⁵ e ressalta a necessidade de seleção criteriosa do controle. O *biocreep* pode ocorrer, por exemplo, quando se utiliza um genérico como controle de um novo produto, ainda que esse genérico tenha sido considerado não-inferior ao produto original. Parece óbvio que A sendo não-inferior a B e C não-inferior a A não significa que C seja não-inferior a B. Em recente publicação¹⁶, foram examinados por meio de simulações os diversos fatores que podem levar ao indesejado *biocreep* – além da escolha do controle, violações do pressuposto de constância foram consideradas fontes de maior potencial para o fenômeno.

População de análise

Os resultados dos estudos clínicos podem ser analisados considerando-se dois possíveis conjuntos de dados. O primeiro inclui toda a população randomizada, independentemente de desistências, perdas ou falta de adesão ao tratamento, e é definido pela intenção de tratar, referido pela sigla ITT (*intention-to-treat*). O segundo conjunto inclui somente os pacientes que completaram o tratamento e que não cometeram nenhuma violação de protocolo; é um sub-conjunto do anterior, referido pela sigla PP (*per-protocol*). Algumas vezes, utiliza-se um conjunto intermediário, conhecido como MITT (*modified intention-to-treat*), que inclui todos os pacientes, exceto aqueles que não receberam nenhuma dose do tratamento.

Para os estudos de superioridade existe o consenso de que o conjunto ITT deve ser preferido. A justificativa é que

essa estratégia previne viés de atrito, preserva a randomização inicial e, mais importante, representa uma salvaguarda contra suspeitas de exclusão consciente ou inconsciente de dados indesejáveis. É também reconhecido que a análise segundo o ITT fornece um resultado mais conservador, diminuindo as diferenças entre tratamentos e, por essa mesma razão, introduz um viés oposto aos estudos de equivalência ou não-inferioridade. A análise segundo o princípio PP, por ser feita em uma população de pacientes mais aderentes ao tratamento, tem maior eficiência em discriminar tratamentos e por isso é, segundo muitos, a estratégia a ser seguida na análise de estudos de não-inferioridade. Contudo, essa recomendação não é livre de controvérsias por inflacionar o erro tipo II. A solução pragmática encontrada pelas principais autoridades sanitárias é recomendar a análise pelos dois métodos e requerer que os resultados sejam robustos e consistentes. O planejamento de análise segundo PP tem implicações no cálculo do tamanho da amostra e deve-se permitir uma provisão adequada para uma taxa estimada de perdas.

Outra condição a ser considerada é a qualidade de execução do estudo. Uma má aderência, medidas imprecisas e processos degradados aumentam a variabilidade e mascaram as diferenças entre tratamentos. O que pode ser devastador para um estudo de superioridade produz efeito inverso nos estudos de não-inferioridade. Diluir diferenças aumenta a probabilidade de resultados falso-positivos, declarando não-inferiores tratamentos que de outra forma não o seriam. A ausência desse “incentivo de qualidade” torna os estudos de não-inferioridade muito mais complexos.

Justificativa ética

O princípio de equipolência clínica (*clinical equipoise*) fornece uma justificativa ética para os estudos clínicos randomizados. Equipolência clínica é o estado de incerteza genuína sobre qual dentre dois ou mais tratamentos é o mais eficaz e seguro – um estudo clínico é conduzido para reduzir esse estado de incerteza. Existe um considerável desacordo sobre o uso de placebo. Situações existem nas quais o seu uso é perfeitamente justificável e outras nas quais o seu emprego não é considerado ético. A última revisão da Declaração de Helsinque (2008)¹⁷ introduziu maior flexibilidade para o uso de placebo em pesquisa clínica, mas ainda persiste muita controvérsia. No Brasil, há uma resolução do Conselho Nacional de Saúde indicando que “os benefícios, riscos, dificuldades e efetividade de um novo método devem ser testados comparando-os com os melhores métodos atuais”¹⁸. Consoante, os estudos de não-inferioridade

representam uma contribuição metodológica para reduzir a exposição de pacientes a placebo. No entanto, a própria definição desses estudos implica que o tratamento em avaliação pode ser, a menos de uma margem, inferior ao controle e isso é, por si só, suficiente para desencadear questionamentos éticos. Uma radical objeção ética aos estudos de não-inferioridade foi recentemente publicada¹⁹ a ponto de sugerir o seu banimento. Seguramente, essa não parece ser a opinião da comunidade científica.

Alguns critérios podem ajudar no escrutínio ético dos estudos de equivalência e não-inferioridade: a condição não deve representar risco de dano sério ou irreversível, existência de incontestáveis benefícios marginais do tratamento em estudo (efeitos colaterais, custo, facilidade de administração etc.), consentimento informado muito bem esclarecido e, muito importante, um controle de qualidade do planejamento e execução do estudo (definição de margem, sensibilidade de ensaio, adequada definição do controle etc.).

Vários são os problemas e desafios para a condução de estudos de não-inferioridade. Além de um grande número de publicações recentes em revistas especializadas, algumas diretrizes podem ser consultadas, como ICH-E9²⁰, ICH-E10²¹, EMEA^{22,23}. Até recentemente, as recomendações do FDA ou eram genéricas, baseadas nas recomendações do ICH, ou demasiado específicas dirigidas a certos produtos, indicações ou áreas terapêuticas determinadas. Entretanto, em março de 2010 o FDA publicou nova e abrangente diretriz²⁴ para o planejamento e análise de estudos de não-inferioridade. Esse documento, embora ainda na sua versão não definitiva, fornece importantes orientações. Para elaborar ou revisar relatórios de estudos de não-inferioridade, alguns checklists são disponíveis. Dois desses são especialmente úteis, um elaborado pelo grupo CONSORT²⁵ e outro pelo *Canadian Partnership Against Cancer* da Universidade de McMaster²⁶. No Brasil, ainda existe um considerável grau de desconhecimento sobre o assunto, sendo frequente o uso de metodologias inapropriadas na condução de estudos de não-inferioridade. Muitos tratamentos declarados não-inferiores não o seriam se os estudos fossem realizados corretamente. Trazer esse tema à discussão mais ampla e estimular o desenvolvimento de guias e diretrizes locais parecem oportunidades a serem aproveitadas.

Referências

1. Fletcher RH. Evaluation of interventions. *J Clin Epidemiol*. 2002;55:1183-90.
2. Altman DG, Bland JM. Treatment allocation by minimisation. *BMJ*. 2005;330:843.
3. Chow SC, Chang M. Adaptive design methods in clinical trials – a review. *Orphanet J Rare Dis*. 2008;3:11.
4. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995;311:485.
5. Alderson P. Absence of evidence is not evidence of absence. *BMJ*. 2004;328:476-7.
6. Evans SR. Noninferiority clinical trials. *Chance*. 2009;22:53-8.
7. Pocock SJ. The pros and cons of noninferiority trials. *Fundam Clin Pharmacol*. 2003;17:483-90.
8. D'Agostino RB Sr, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues – the encounters of academic consultants in statistics. *Stat Med*. 2003;22:169-86.
9. Kaul S, Diamond GA, Weintraub WS. Trials and tribulations of non-inferiority: the ximelagatran experience. *J Am Coll Cardiol*. 2005;46:1986-95.
10. Chow SC, Shao J, Wang H. Sample size calculations in clinical research. 2nd Ed. Boca Raton: Chapman & Hall/CRC Biostatistical series; 2008.
11. Lehr R. Sixteen S-squared over D-squared: a relation for crude sample size estimates. *Stat Med*. 1992;11:1099-102.
12. Pater C. Equivalence and noninferiority trials – are they viable alternatives for registration of new drugs? (III). *Curr Control Trials Cardiovasc Med*. 2004;17:8.
13. Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med*. 2006;145:62-9.
14. James Hung, HM, Wang SJ, Tsong Y, Lawrence J, O'Neil RT. Some fundamental issues with non-inferiority testing in active controlled trials. *Stat Med*. 2003;22:213-25.
15. Fueglistaler P, Adamina M, Guller U. Non-inferiority trials in surgical oncology. *Ann Surg Oncol*. 2007;14:1532-9.
16. Everson-Stuart SP, Emerson SS. Bio-creep in non-inferiority clinical trials [Internet]. Biostatistics Working Paper Series, University of Washington, Seattle; 2010 [citado 2010 ago 13]. <http://www.be-press.com/uwbiostat/paper359>
17. World Medical Association. Ethical principles for medical research involving human subjects, 59th WMA General Assembly, Seoul [Internet]. 2008. [citado 2010 ago 13]. <http://www.wma.net/en/30publications/10policies/b3/index.html>
18. Brasil. Ministério da Saúde. Conselho Nacional de Saúde. Resolução 404/2008 de 1 de Agosto de 2008. Brasília: Diário Oficial da União – DOU 186 de 25 de Setembro de 2008; 2008. p. 45 ISSN 1677-7042.
19. Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet*. 2007;370:1875-7.
20. International Conference of Harmonization (ICH). Guideline E9 – Statistical principles for clinical trials [Internet]. 1998. [citado 2010 ago 16]. <http://www.ich.org/cache/compo/475-272-1.html#E9>
21. International Conference of Harmonization (ICH). Guideline E10 – Choice of control groups and related design and conduct issues in clinical trials [Internet]. 2000 [citado 2010 ago 16]. <http://www.ich.org/cache/compo/475-272-1.html#E10>

22. Committee for Proprietary Medicinal Products. Points to consider on switching between superiority and non-inferiority. *Br J Clin Pharmacol*. 2001;52:223-8.
23. Committee for Proprietary Medicinal Products for Human Use. Guideline on the choice of the non-inferiority margin [Internet]. 2005. [citado 2010 ago 16]. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf
24. FDA Guidance for Industry. Non-inferiority clinical trials: draft guidance. US Dept of Health and Human Service FDA/CDER/CBER [Internet]. 2010. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf>
25. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA*. 2006;295:1152-60.
26. Canadian Partnership Against Cancer. Capacity Enhancement Program: non-inferiority and equivalence trials checklist [Internet]. McMaster University; 2009. [citado 2010 ago 13]. <http://fhs.mcmaster.ca/cep/documents/CEPNon-inferiorityandEquivalenceTrialChecklist.pdf>

Correspondência:

Valdair Ferreira Pinto
Alameda dos Aicás, 229
CEP 04086-000 – São Paulo, SP
E-mail: valdairpinto@hotmail.com