

Entendendo Ensaio Clínicos Randomizados

Understanding Randomized Controlled Trials

MARCO AURÉLIO PINHO DE OLIVEIRA¹; RAPHAEL CÂMARA MEDEIROS PARENTE²

¹ Doutor em Epidemiologia pelo Instituto de Medicina Social da Universidade do Estado do Rio de Janeiro. Chefe do Setor de Ginecologia da UERJ; ² Doutor em Ginecologia pela UNIFESP (Reprodução Humana). Mestre em Epidemiologia pelo Instituto de Medicina Social da Universidade do Estado do Rio de Janeiro.

Bras. J. Video-Sur, 2010, v. 3, n. 4: 176-180

Accepted after revision: Julho, 2010.

Os ensaios clínicos constituem-se numa poderosa ferramenta para a avaliação de intervenções para a saúde, sejam elas medicamentosas ou não¹. O primeiro ensaio clínico, nos moldes que hoje conhecemos, foi publicado no final da década de 40², quando o estatístico Sir Austin Bradford Hill alocou aleatoriamente pacientes com tuberculose pulmonar em dois grupos: os que receberiam estreptomina e os que não receberiam o medicamento. Desta forma, ele pode avaliar, de maneira não viesada, a eficácia deste medicamento. Em que pese a publicação crescente de ensaios clínicos controlados, alguns aspectos do desenho e da análise ainda são mal compreendidos e interpretados de forma equivocada.

Os ensaios clínicos são estudos onde um grupo de interesse em que se faz uso de uma terapia ou exposição é acompanhado comparando-se com um grupo controle. Diferente dos estudos observacionais em que o pesquisador não interfere na exposição, nesse estudo o pesquisador planeja e intervém ativamente nos fatores que influenciam a amostra, minimizando assim a influência dos fatores de confundimento. A alocação dos sujeitos de pesquisa pode ser de forma aleatória (randomizada) ou não-aleatória.

Embora os ensaios clínicos randomizados (ECRs) sejam o padrão-ouro para determinação de efeito de uma terapêutica, por diversos motivos, seu uso na cirurgia não é tão difundido como nas outras especialidades médicas³. Ensaios clínicos são laboriosos e custosos em realizá-los e na cirurgia trazem consigo diversos desafios práticos e metodológicos na sua elaboração e realização.

As comparações randomizadas de intervenções cirúrgicas têm sido realizadas por muitos anos. Várias intervenções cirúrgicas usadas por períodos

longo de tempo têm-se mostrado ineficazes e depois descartadas após submetidas a esses estudos⁴. A ligadura da artéria mamária interna foi descartada após resultados de dois pequenos ensaios clínicos não terem demonstrado diferença em relação ao placebo. Mas surpreendentemente, o número de ensaios clínicos na cirurgia não acompanha o de outras especialidades.

Uma revisão encontrou um aumento no número de ensaios clínicos randomizados de 1990 a 2000⁵. No entanto, apenas 3,4% do total de artigos em importantes revistas cirúrgicas foram ensaios clínicos⁶. O contraste com outras áreas podem ser vistas por comparações entre auditorias de bases de dados para a prática clínica em medicina interna e diversas especialidades cirúrgicas. Considerando que metade das intervenções em medicina interna foram julgadas com base em evidências de ensaios clínicos, duas auditorias cirúrgicas relataram um quarto ou menos quando avaliadas evidências cirúrgicas⁶⁻⁷.

O principal problema relacionado aos ensaios clínicos na cirurgia e que dificultam sua confecção e realização é a dificuldade no mascaramento. O mascaramento é o processo de retenção da informação sobre as intervenções atribuídas a cada grupo e é um elemento-chave na concepção de ECRs. Quando feito com sucesso, é reconhecido como tendo um papel importante na prevenção da introdução de um viés de informação. O cegamento pode ser aplicado aos participantes, aos pesquisadores e aos avaliadores dos resultados do estudo. Em ensaios clínicos cirúrgicos, o mascaramento coloca um desafio maior do que para terapias medicamentosas. É impossível impedir que o cirurgião saiba qual técnica ele estará usando. Logo, é impossível o mascaramento do médico. Os avaliadores podem, e devem, preferencialmente, ser mas-

carados quanto aos resultados. No entanto, o mascaramento não ocorre geralmente em intervenções cirúrgicas para nenhum dos participantes do estudo. Isso ocorre devido ao hábito do paciente retornar ao médico que o operou para o seguimento. O mascaramento dos pacientes pode ser conseguido quando se usa a mesma via para cirurgias diferentes. O mesmo é mais difícil de se obter quando as vias são diferentes como por exemplo laparotômica comparada com laparoscópica e virtualmente impossível quando se compara cirurgia com medicamentos. Embora cirurgia como placebo seja teoricamente possível, motivos éticos impedem sua concepção na prática, embora haja relatos na literatura⁸.

Ensaio clínico bem conduzido é o padrão-ouro entre os estudos para guiarem a nossa prática clínica diária. A elaboração de um ECR visa a um estudo que seja suficientemente grande para que um efeito clínico importante seja estatisticamente significativo, mas não tão grande para que desfechos pouco importantes sejam significativos. Além disso, quando excessivamente grandes, os ECRs podem ser proibitivamente caros⁹. O tamanho da amostra para qualquer estudo é calculado com base na quantidade dos erros tipo I e II que o investigador está disposto a tolerar. Um ponto forte dos ECRs é que a alocação aleatória evita o fator de confundimento por dispersar igualmente esses fatores entre os grupos de forma paritária. No entanto, existem importantes inconvenientes de um único ECR de grande porte. Primeiro, alguns ECRs são tão rigorosos na sua metodologia que os resultados podem somente servir para essa população estudada (problemas de validade externa ou de generalização). Além disso, os pacientes que participam de ensaios clínicos podem ter uma linha de base de risco para o desfecho de interesse muito maior do que a média dos pacientes da população ou podem ser oriundos de populações muito diferentes. Por exemplo, alguns dos estudos que estudaram a ingestão de cálcio para a prevenção da pré-eclâmpsia (embora com metodologia rigorosa) foram realizados em países com dietas muito diferentes entre si o que impede conclusões para populações com dietas outras. Isso pode impedir uma generalização dos resultados.

Além disso, alguns ensaios clínicos podem não possuir um número suficiente de indivíduos em subgrupos importantes. Uma preocupação final com o ECR é a questão do viés de publicação. O viés de publicação ocorre quando os resultados dos estudos que relatam resultados positivos (com significância

estatística) são mais prováveis de serem publicados do que estudos com resultados negativos. Os estudos que relatam resultados significativos podem resultar em um maior número de publicações e também poderão ser publicados em jornais de maior impacto. Isso pode ter impacto na confecção de metanálises que podem não englobar ensaios com resultados desfavoráveis.

Um ensaio clínico randomizado é um estudo prospectivo em humanos comparando o efeito e o valor de uma intervenção contra um controle¹⁰.

1. Ensaio clínico randomizado deve ser considerado quando:

- a. Há incerteza sobre o efeito de uma exposição ou tratamento;
- b. A exposição pode ser modificada no estudo.

2. Potenciais limitações de ensaios clínicos randomizados incluem:

- a. Limitada generalização da população em estudo.
- b. Limitada generalização do ambiente em estudo.
- c. Ensaio clínico randomizado responde a uma questão de estudo específica.

3. As medidas de magnitude de efeito usuais em ensaios clínicos randomizados são o risco relativo e a diferença de risco.

Em ECRs, um tratamento ou procedimento pode ser comparado a nenhum tratamento, a uma terapêutica semelhante (mesma classe de drogas nos *head-to-head trials*), com um placebo ou a uma norma preexistente de acompanhamento¹⁰. ECRs são realizados sob condições controladas para garantir que a hipótese do estudo é testada de uma forma reprodutível. Protocolos de ensaios clínicos randomizados muitas vezes não podem ser repetidos na prática clínica. O ambiente de estudo especializado de um estudo randomizado pode resultar em conclusões válidas internamente, mas os resultados não podem ser generalizados para pacientes com a mesma condição fora do estudo¹⁰.

4. Ensaio clínico randomizado são projetados para responder definitivamente uma questão específica, concentrando-se estritamente em uma hipótese de pesquisa, isolando o efeito de uma ou de

um pequeno número de terapias. Ensaios clínicos randomizados não são projetados para avaliar os mecanismos pelos quais uma terapia pode produzir benefícios ou prejuízos.

Costumam ter uma aplicabilidade clínica limitada. Conclusões retiradas de ECRs são limitadas a situações em que a exposição de interesse pode ser modificada em um ambiente de estudo, por exemplo, uso de medicamentos ou mudanças de estilo de vida. Há inúmeras exposições que são altamente relevantes para estudo, mas não podem ser facilmente modificadas, tais como genes, marcadores séricos, condições socioeconômicas etc.. Ensaios clínicos randomizados são limitados a situações clínicas específicas nas quais a exposição de interesse pode ser facilmente modificada. Também não podem ser usados em situações nas quais a exposição seja reconhecidamente deletéria, por não ser ético. Por exemplo, não podemos fazer com que alguém fume para compararmos com um grupo que não fume por já sabermos dos malefícios do fumo.

A pesquisa de drogas pela indústria envolve ECRs nas suas diferentes fases. São realizadas basicamente em 4 fases¹⁰:

Estudos Fase I

O primeiro passo no desenvolvimento de uma nova droga é entender se o medicamento é bem tolerado em um pequeno número de pessoas. Embora não seja um ensaio clínico, estes tipos de estudos são referidos como estudos de fase I. Os participantes de estudos de fase I são adultos saudáveis ou pessoas com a doença específica que a droga se destina a tratar. Ocasionalmente, estudos de fase I não podem ser realizados em adultos saudáveis porque a droga tem inaceitáveis efeitos adversos, tais como agentes quimioterápicos. Estudos de fase I procuram determinar até que dose uma droga pode ser administrada antes de ocorrer toxicidade inaceitável. Esses estudos são iniciados com baixas doses em número limitado de pessoas e, em seguida, aumenta-se a dose gradualmente.

Estudos de Fase II

São projetados para avaliar se um medicamento possui atividade biológica e para determinar sua segurança e tolerabilidade.

Estudos de Fase III / IV

Os estudos de fase III são ensaios clínicos randomizados delineados para avaliar a eficácia e a

segurança de uma intervenção. Os resultados dos estudos de fase III são desfechos clínicos, tais como morte ou sobrevida livre de tumor. As avaliações de segurança ocorrem durante um período mais longo quando comparados com estudos de fase II. Estudos de fase IV ocorrem após aprovação e avaliam os resultados associados a uma droga ou intervenção na prática clínica com o uso na população geral.

Na avaliação crítica de um ensaio clínico randomizado, devemos fazer algumas perguntas importantes, usando algumas informações já assinaladas anteriormente.

A primeira pergunta que se deve fazer é “Os pacientes selecionados neste estudo são semelhantes aos pacientes que eu trato?”. O trabalho deve dizer claramente quais foram os critérios de inclusão e exclusão (idade, sexo, doenças anteriores, etc..). Trabalhos com critérios muito rígidos de inclusão são mais restritivos na generalização das conclusões (porém são mais específicos). Deve mencionar ainda como os pacientes foram selecionados da população geral (Só entraram os pacientes que procuraram o hospital terciário? São apenas voluntários remunerados? ou são apenas os pacientes de uma comunidade carente próxima ao hospital?). Lembrar que só podemos generalizar para as populações semelhantes ao do estudo em questão. Cabe ao leitor decidir se as diferenças entre a população selecionada no estudo e a população que lhe interessa são importantes ou não.

Um dos aspectos fundamentais é o evento final de interesse. Muitos trabalhos são até metodologicamente bem montados, porém estudam variáveis que não são muito úteis clinicamente, ou que não representam aquilo que deveriam. Por exemplo, no tratamento da endometriose com uma medicação antiestrogênica, o pesquisador verifica somente uma diminuição estatisticamente significativa do score da AFS (quantifica a extensão da doença) e conclui que o tratamento é eficaz. Porém, o que a endometriose causa é basicamente infertilidade e dor pélvica, e o estudo em questão não avaliou o que realmente interessa (evento final de interesse), ou seja, se a paciente melhorou da dor ou se conseguiu engravidar (não interessa à paciente melhorar somente o score da AFS). Outro exemplo seria uma pesquisa para verificar a eficácia de um novo quimioterápico, tendo como parâmetro apenas a diminuição da massa tumoral (porém, a diminuição do tumor pode não se correlacionar com a sobrevida, variável de interesse para a paciente).

Outro aspecto importante é saber como foi feita a distribuição dos indivíduos entre o grupo tratamento e o grupo controle (ou placebo). Trabalhos que não incluem grupo controle ou placebo perdem muito (correm o risco de mostrar que um tratamento é eficaz quando na verdade não é), pois o placebo pode produzir efeitos (p.ex. melhora da cefaléia) superiores a 50% (pode inclusive ser melhor que o medicamento estudado!), ou seja, a melhora do paciente pode acontecer não pelo tratamento em questão, mas sim por outros motivos (história natural da doença, aspectos psicológicos etc.). Algumas vezes a utilização de um placebo (não tratamento) é inadequado, como por exemplo no tratamento de uma doença inflamatória pélvica aguda. Nestes casos, o pesquisador pode comparar um novo medicamento usando como grupo controle o tratamento considerado padrão.

O próximo passo a ser verificado é saber como foi feita a randomização entre os grupos tratamento e controle. A randomização implica na distribuição aleatória dos pacientes, de modo que cada caso possa ter a mesma chance de ser alocado nos diferentes grupos. Estudos controlados (i.e., com grupo controle) não randomizados perdem na qualidade da evidência científica, pois na maioria das vezes os grupos não são homogêneos (grupos viciados). Por exemplo, num estudo comparativo entre a cirurgia de Burch laparoscópica *versus* laparotômica, pode ser que sejam selecionadas para a laparoscopia pacientes mais magras e com queixas mais leves de incontinência urinária de esforço. Deste modo, não se pode comparar estes dois diferentes grupos de pacientes (neste caso desfavorece a cirurgia por laparotomia). A randomização acaba ajustando esses fatores conhecidos (idade, peso, grau da doença etc.) assim com fatores desconhecidos (p.ex. algum fator genético desconhecido e que possa alterar os resultados). A randomização adequada deve ser feita com números aleatórios gerados por computador (também existem tabelas apropriadas) que são colocados em envelopes lacrados numerados que são abertos imediatamente antes da aplicação do tratamento. Métodos não adequados de randomização (quasi-randomizados) incluem, por exemplo, fazer cirurgia por laparoscopia 2^{as} e 4^{as} feiras e por laparotomia 3^{as} e 5^{as} feiras, ou então fazer uma cirurgia por laparoscopia e outra por laparotomia (alternando as vias). Estes métodos propiciam falhas na distribuição aleatória, mesmo que inconscientemente. Por exemplo, o cirurgião (ou quem encaminhou a paciente para a cirurgia e que conhece

o esquema do estudo) pode achar a paciente muito obesa para a laparoscopia e subliminarmente (ou propositalmente?) adiar a cirurgia por um dia, coincidindo com a vez da laparotomia. A randomização adequada evita este tipo de erro na maioria das vezes. Entretanto, a randomização bem feita não garante a homogeneidade dos grupos (principalmente com amostras pequenas). É de bom tom verificar antes da análise dos dados se houve realmente uma distribuição semelhante dos pacientes (particularmente das características que influenciam diretamente no resultado, como: estágio da doença, idade, peso, etc.). Grupos não-balanceados podem enviesar os resultados.

Deve-se também estar atento se o método utilizado na mensuração do evento final de interesse foi o mascaramento duplo (duplo-cego). Isto implica que nem o paciente e nem o profissional que está fazendo a avaliação (p.ex. verificar o grau de dor no pós-operatório em pacientes que usaram ou não marcaína na ferida operatória) sabem qual tratamento foi aplicado. O médico pode, inconscientemente, não valorizar as queixas das pacientes que pertencem ao grupo do novo tratamento, falseando os resultados. O próprio paciente pode ser simpático a um determinado tipo de tratamento e não informar ao médico todas as queixas que tem. O método duplo-cego tenta eliminar este potencial tipo de erro. Nem sempre existe a possibilidade de se conduzir um estudo duplo-cego (comparação entre tratamento clínico *versus* cirúrgico), porém não se deve deixar de utilizá-lo quando possível.

Outro aspecto importante é o tempo de seguimento (follow-up). Em primeiro lugar é preciso certificar-se que o tempo de seguimento foi o suficiente para avaliar o evento final de interesse. Por exemplo, na transecção laparoscópica dos ligamentos útero-sacros, com o objetivo de aliviar a dismenorréia, o tempo de seguimento deve ser pelo menos superior a 1 ano, pois sabe-se do número elevado de recidivas após este período. Outro aspecto relevante é verificar se o percentual de pacientes que abandonaram o seguimento foi semelhante entre os diversos grupos. Taxa de seguimento inferior a 70% praticamente inviabiliza a análise dos dados, pois o abandono pode ter acontecido por efeitos colaterais ou até pelo óbito decorrente do tratamento.

Os erros mencionados anteriormente são sistemáticos, diferente dos erros aleatórios, que podem acontecer quando trabalhamos com qualquer amostra. A análise estatística adequada visa identificar se

as diferenças obtidas nos resultados foram por conta dos erros aleatórios ou se a diferença realmente existe entre os tratamentos aplicados. Os conceitos básicos da estatística serão abordados no próximo volume da revista.

LEITURAS SUPLEMENTARES

- 1) Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. A Medical Research Council Investigation. *BMJ*. 1948;2:769-82.
- 2) Coutinho ESF, da Cunha GM. Conceitos básicos de epidemiologia e estatística para a leitura de ensaios clínicos controlados. *Rev Bras Psiquiatr*. 2005;27(2):146-51.
- 3) Cook J. The challenges faced in the design, conduct and analysis of surgical RCTs. *Trials* 2009, 10:9.
- 4) Ko CY, Sack J, Chang JT, Fink A: Reporting randomised, controlled trials: where quality of reporting may be improved. *Dis Colon Rectum* 2002, 45:443-447.
- 5) Wente MN, Seiler CM, Uhl W, Büchler MW: Perspectives of evidence-based surgery. *Dig Surg* 2003, 20:263-269.
- 6) Ellis J, Mulligan I, Rowe J, Sackett DL: Inpatient general medicine is evidence based. *Lancet* 1995, 364:407-410.
- 7) Howes N, Chagla L, Thorpe M, McCulloch P: Surgical practice is evidence based. *Br J Surg* 1997, 84:1220-1223.
- 8) Kenny SE, Shankar KR, Rintula R, Lamont GL, Lloyd DA: Evidence-based surgery: interventions in a regional paediatric surgical unit. *Arch Dis Child* 1997, 76:50-53.
- 9) Scifres CM, Iams JD, Klebanoff M, Macones GA. Metaanalysis vs large clinical trials: which should guide our management? *Am J Obstet Gynecol* 2009;200:484.e1-e5.
- 10) Kestenbaum B. *Epidemiology and Biostatistics: An Introduction to Clinical Research*, DOI 10.1007/978-0-387-88433-2_7, © Springer Science Business Media, LLC 2009.

Endereço para Correspondência:

MARCO AURÉLIO PINHO DE OLIVEIRA
Rua Coelho Neto, 55 / 201
Tel.: (21) 9987-5843
E-mail: maurelio@infolink.com.br