

Será que ensaios clínicos aleatórios podem inspirar decisões clínicas sobre pacientes individuais?*

David Mant

Clínicos gerais não querem ser curandeiros. Eles concordam com Richard Asher (1961) que o uso da massa de pó de peixe fossilizado para tratar erisipela é pura conversa fiada. Eles concordam sem hesitar que a clínica médica, incluindo os remédios transmitidos pelo folclore clínico, deve estar sujeita à avaliação empírica. Eles reconhecem a importância dos ensaios clínicos aleatórios na avaliação da eficiência terapêutica e participam ativamente da Cochrane Collaboration¹ (Silagy, 1993). Mas, como apontou Hart, em 1997, em sua conferência Cochrane, os clínicos gerais preocupam-se cada vez mais com a aplicação grosseira de evidências obtidas por meio de ensaios randômicos à prática clínica. Uma autoridade do porte do presidente do Colégio Real de Clínicos Gerais do Reino Unido foi recentemente co-autor de um ensaio no qual descreveu a medicina baseada em evidências como a “nova divindade na clínica médica”, acusando-a de diminuir a importância das relações humanas por ignorar o fato de que decisões clínicas no atendimento primário envolvem a consideração de problemas e preocupações singulares de cada paciente (Sweeney et al., 1998).

* Originalmente publicado em *The Lancet*, vol. 353, 27 de fevereiro de 1999. Tradução de Monica Seincman.

1. Rede internacional de pesquisadores que, voluntariamente, organizaram-se para produzir revisões sistemáticas dos resultados de ensaios clínicos aleatórios, criada em 1992. Já realizou 795 artigos de revisão. Quinze centros Cochrane preparam outros 738. Archibald Cochrane (1909-1988), epidemiologista inglês, foi seu idealizador. Autor de *Effectiveness and Efficiency. Random Reflections on Health Services* (London: Nuffield Provincial Hospitals Trust, 1972). (Nota dos Editores)

O que nos contam os ensaios clínicos sobre a prática clínica com pacientes individuais? Tais experimentos fornecem um prognóstico estatístico do efeito provável de uma dada intervenção, geralmente baseado no resultado “médio” agregado de todos os participantes do experimento. Este prognóstico é freqüentemente expresso como o chamado *número necessário para tratamento* (NNT)² para conseguir uma unidade de benefício, tal como a evitação de uma morte ou o ganho de um ano extra de vida; da mesma maneira, resultados adversos são expressos como o *número necessário para o dano* (NNH)³. Para os clínicos, a questão é saber se e como o NNT e o NNH podem ser aplicados aos pacientes que os procuram. Esta preocupação suscita duas questões. Primeira: serão os resultados dos ensaios generalizáveis, de tal modo a se esperar que a resposta de um paciente em particular a uma dada intervenção seja similar à resposta dos participantes do experimento à mesma intervenção? Segunda: é possível refinar o prognóstico estatístico baseado no resultado “médio” do experimento para obter um prognóstico mais preciso do resultado no caso particular de um dado paciente? Vale a pena examinar ambas as questões detalhadamente.

Generalizando a partir da evidência do experimento

Participantes de ensaios clínicos raramente representam o conjunto da população. Em geral, os participantes são mais saudáveis, mais jovens, e de nível social mais elevado do que as pessoas às quais a intervenção se destina. Será que isto importa? Pode ser que em alguns casos não. Há muitas situações de rotina na prática médica, da cera do ouvido à verruga plantar, em que resultados de ensaios clínicos podem ser aplicados a um amplo espectro da população, pois as condições estão bem definidas e as intervenções avaliadas são simples. Glasziou e Irwig (1995) afirmam que esta situação é norma e não exceção e citam evidências de que tanto o dano como o benefício relativo das intervenções mostram-se constantes em subgrupos das populações testadas. Fahey (1998) também aponta que o resultado principal do fato de os experimentos falharem em recrutar pacientes mais velhos, e portadores de comorbidades – que representariam melhor aqueles que se apresentam aos clínicos gerais – tem sido o de subestimar tanto o benefício relativo quanto absoluto das intervenções sob avaliação.

No entanto, há exemplos claros em que o caráter atípico dos participantes de ensaios clínicos é relevante. Evidências advindas da meta-análise de ensaios realiza-

2. *Number Needed for Treat.* número de indivíduos que um clínico precisaria tratar para evitar um resultado adverso. A medida estabelece o benefício de uma intervenção comparada a nenhuma intervenção num grupo de pessoas sob risco de apresentar a doença, na situação em que eventos adversos ainda seriam esperados mesmo com a intervenção. (Nota dos Editores)
3. Em inglês, *Number Needed for Harm.*

dos para avaliar o tratamento com antibióticos de crianças com otite média aguda sugerem que o benefício é duvidoso (Del Mar et al., 1997); no entanto, considera-se discutível a aplicabilidade deste resultado à prática clínica – por causa das críticas de que o recrutamento exclui as crianças mais doentes (Burke et al., 1997; Fromm et al., 1990). Há uma preocupação similar entre os clínicos gerais em relação aos experimentos sobre o uso de anticoagulantes nos casos de fibrilação auricular. A questão, aqui, não é tanto a avaliação do benefício (o NNT para a prevenção do derrame), mas o potencial de dano (o NNH para hemorragia maior). Glasziou e Irwig (1995) usaram a avaliação de dano nestes experimentos como um modelo do chamado efeito fixo⁴. No entanto, a maioria dos experimentos de fibrilação auricular excluiu pacientes com problemas sociais ou comorbidade significativa – a exclusão no experimento SPINAF⁵ foi superior a 90% (Sweeney et al., 1995). A exclusão possivelmente melhorou a qualidade metodológica e a precisão do conjunto de ensaios (por diminuir a não adesão ao protocolo e as falhas de acompanhamento); no entanto, ela também explica porque, nos ensaios, os índices de hemorragia foram substancialmente mais baixos do que os estudos de observação de pacientes tratados com anticoagulantes na prática clínica de rotina (Fihn et al., 1996). Além disso, Rothwell (1995) mostrou em relação ao Experimento de Aspirina do Reino Unido ao Experimento Europeu de Cirurgia da Carótida que o resultado dos pacientes é influenciado não apenas pelo risco basal, mas também pela variação no efeito do tratamento em relação ao risco basal.

O exemplo da fibrilação auricular também revela a dificuldade de generalizar a evidência do experimento quando a intervenção avaliada é complexa. O tratamento com anticoagulantes não é uma intervenção simples. Requer um paciente motivado e organizado, apoiado por um serviço de saúde local motivado, organizado e tecnicamente competente. A frequência de hemorragia em um paciente tratado com warfarina reflete não apenas a comorbidade do paciente e sua adesão ao protocolo, mas também a qualidade do apoio médico e de enfermagem. O uso da informação sobre o serviço local a fim de refinar um prognóstico requer que o clínico julgue se o serviço oferece ao paciente cuidados de qualidade similar àqueles oferecidos aos que participaram do ensaio clínico. Infelizmente, muitas intervenções do atendimento primário em serviços de saúde são ainda mais complexas do que o tratamento com anticoagulantes. Por exemplo, questiona-se a aplicabilidade dos dois maiores ensaios sobre a prevenção de doença cardiovascular realizados no Reino Unido durante os 10

4. No artigo citado, os autores escrevem que “pacientes sob maior risco terão o maior benefício absoluto”; e que “o benefício para os pacientes em geral aumenta com o risco enquanto o dano permanece comparativamente fixo” (Nota dos Editores).

5. Sigla para *Stroke Prevention in non Rheumathic Atrial Fibrillation*. Estudo realizado entre 1987 e 1990, em 16 hospitais norte-americanos de veteranos. Engajou 525 voluntários, 260 dos quais tratados com warfarina contra 265 recebendo placebo.

últimos anos, uma vez que a abordagem “caixa-preta” não definiu adequadamente a intervenção em termos da experiência dos pacientes e enfermeiras e do contexto social da intervenção. Este problema pode ser contornado em certa medida pela realização de pesquisa qualitativa paralela durante ensaios clínicos randômicos sobre intervenções complexas (Bradley et al., prelo). Mas não pode ser eliminado.

A precisão do prognóstico individual

Se é fácil reproduzir a intervenção, e se um paciente em particular for suficientemente parecido com os que participaram do experimento a ponto de uma diferença no resultado parecer improvável, seria esta a solução para o problema da generalização? Não para o clínico, porque a maioria dos experimentos clínicos exibe substancial heterogeneidade de efeito para pacientes individuais. Poucos pacientes ficam satisfeitos de serem informados de que a probabilidade de um resultado positivo é de 20% no grupo de intervenção e de 5% no grupo de controle ou que o NNT de um determinado ensaio é 7. O que eles querem saber é se estarão entre aqueles 20% para os quais a intervenção funcionou.

Um experimento clínico responderá a este tipo de pergunta somente se tiver sido concebido para fornecer informações sobre aquelas características dos pacientes capazes de influir no resultado. Uma destas características é a preferência por um tratamento, que não é simplesmente uma questão de escolha de tratamento, mas sim aquilo que McPherson e colegas (1997) denominaram “o efeito terapêutico da preferência do paciente”. Este efeito é, em algumas circunstâncias, similar ao efeito placebo (O’Boyle, 1993) e pode ter magnitude suficiente para fazer uma intervenção parecer efetiva mesmo que não traga benefício fisiológico adicional, particularmente em ensaios clínicos nos quais é difícil manter a condição de o paciente ignorar se está engajado no grupo de controle ou no grupo que recebe intervenção. Contudo, é difícil identificar o efeito da preferência do paciente no contexto de um experimento clínico (Rucker, 1989; Brewin & Bradley, 1989) (...) Outros efeitos de segunda ordem, que podem ser importantes na predição do resultado para pacientes individuais, tendem a ficar subordinados à investigação do efeito principal pesquisado.

Uma outra abordagem para o desejo clínico de precisar o prognóstico individual é aplicar o benefício fixo encontrado, em um dado experimento, ao “risco” basal do paciente (Glasziou & Irwig, 1995). Esta abordagem pode, às vezes, envolver o reconhecimento de um padrão complexo como um “juízo clínico”. Mas Glasziou e Irwig (1995) e Fahey (1998) afirmam que os julgamentos clínicos podem ser quantificados tendo como referência estudos sobre grupos de observação. Os exemplos que eles dão focalizam características como idade e medidas fisiológicas de colesterol e pressão sanguínea. Sackett (1996) sugeriu que o clínico poderia aperfeiçoar a sua decisão em casos particulares se ponderasse a evidência experimental

face ao valor relativo que o paciente deposita nos resultados positivo e negativo do ensaio clínico – o que é o reconhecimento, mais uma vez, da importância das percepções e valores dos pacientes na tomada de decisão.

No final, contudo, o refinamento do prognóstico para um paciente individual só é possível na medida que o experimento forneça evidências. Glasziou e Irwig (1995) reconhecem que o argumento deles baseia-se no pressuposto de que o benefício ou o dano relativo das intervenções (ou seja, o efeito do tratamento) não varia substancialmente quando aplicado a populações diferentes. Apesar de ser possível sustentar este pressuposto, ao menos nas intervenções simples, para subgrupos determinados dentro das populações experimentais, a descoberta de Rothwell (1995) sobre a variação do efeito do tratamento com o risco basal já foi citada, e a heterogeneidade do efeito na população geral pode ser substancial. O problema é que ensaios clínicos aleatórios não são a forma ideal para tratar esta heterogeneidade – são mais eficientes quando a heterogeneidade da população experimental é insignificante. As limitações da análise de subgrupos, exceto quando baseadas em estratificação prévia, estão bem documentadas (Oxman & Guyatt, 1992).

Finalmente, Guyatt e colegas (1998) propõem não levar em consideração apenas o efeito médio dentro de um experimento, mas também a distribuição dos efeitos. Os métodos que eles sugerem para analisar esta distribuição provocaram muita discussão (Senn, 1998), mas o ponto principal não está em questão – uma intervenção pode fornecer benefício (ou malefício) clinicamente importante para uma proporção significativa de pacientes, mesmo que o efeito médio não seja nem estatística nem clinicamente significativo.

Implicações para uma futura pesquisa

O paradoxo dos ensaios clínicos é que eles são o melhor modo de avaliar se uma intervenção funciona, mas indiscutivelmente o pior para avaliar quem se beneficiará dela. Uma resposta é pedir mais e maiores experimentos clínicos, estratificados para todas as características que podem conduzir à heterogeneidade de efeitos, de modo que haja evidência experimental disponível para qualquer possível subgrupo de pacientes. No limite, é uma convocação para uma série infinita de experimentos com “n=1”. Esta abordagem é ingênua por três razões: não é factível; não é viável financeiramente; e não fornecerá a evidência de que precisamos. Além disso, o equilíbrio⁶ terapêutico (equivalência de riscos em benefícios entre diferentes tratamentos) não pode ser assumido em todas as áreas de incerteza clínica (Kerridge et al., 1998) e

6. *Equipoise*, requisito de validade científica de ensaios clínicos aleatórios. Se respeitado, garante que só seja engajado num experimento aquele paciente a respeito do qual houver substancial incerteza sobre que tratamento lhe trará o maior benefício.

o ponto de equilíbrio pode ser diferente para o grupo (Lilford & Jackson, 1995). Assim, a tomada de decisão clínica requer evidência de efetividade vinda de experimentos clínicos e evidência de aplicabilidade a pacientes individuais vinda de estudos de observação. O problema é o equilíbrio de investimentos entre as duas abordagens metodológicas.

Para medir a eficiência, a população do experimento deve ser selecionada e estruturada para maximizar a possibilidade de generalização. De alguma forma, esta abordagem é similar à amostragem. A população experimental ótima será obtida por meio de estratificação para as características dos pacientes que tendem a conduzir à heterogeneidade do efeito relativo do tratamento, enquanto tenta maximizar a homogeneidade em cada estrato. Esta abordagem coaduna-se com o apelo de Charlton (1996) por questões altamente focadas para as quais se buscará respostas nos grupos homogêneos de participantes do estudo; sua utilização será facilitada com o aumento do potencial da genética molecular em caracterizar funções metabólicas e prever a resposta individual aos agentes farmacológicos (Bell, 1998). A principal limitação para esta abordagem é a disponibilidade de recursos, o grau de colaboração profissional e a factibilidade de alcançar e sustentar a randomização dos ensaios.

Também deve ser reconhecido explicitamente que, para aplicar os resultados dos experimentos a pacientes individuais, deve haver um investimento paralelo em estudos de observação – tanto quantitativos como qualitativos. Estes estudos são importantes não apenas para fornecer a avaliação do risco basal necessário para modelar o benefício absoluto para um dado paciente (Glasziou & Irwig, 1995; Fahey, 1998; Oxman & Guyatt, 1992), mas também para caracterizar a experiência dos pacientes com a intervenção; a frequência de ocorrências adversas de longo prazo; o efeito no resultado de características como a preferência dos pacientes; o efeito da intervenção em diferentes estágios do curso da doença; e a possibilidade efetiva da implementação da intervenção na prática rotineira.

Ligando a lacuna de evidência clínica

A complexidade do uso das evidências de ensaios clínicos para a tomada de decisões na clínica geral não está em questão. Hart (1997) aponta para a complexidade dos problemas sociais no âmbito dos quais os problemas clínicos surgem e devem ser resolvidos, e conclui que a maior parte da clínica geral depende mais “de interpretar de maneira personalizada os padrões singulares da doença, do que de reconhecer seus padrões generalizados”. Van Weel (1996) relata que mais de 25% dos pacientes com idade superior a 65 anos apresentam comorbidade significativa. Naylor (1995) descreve “zonas cinzas” onde a evidência científica é incompleta ou conflitante e comenta que “o que está preto e branco no abstrato pode rapidamente tornar-se cinza na prática, pois o que os clínicos procuram é encontrar as necessi-

dades de seus pacientes”. Sullivan e colegas (1996) analisam a tarefa da consulta no atendimento primário e concluem que o uso da evidência dos ensaios clínicos é frequentemente subsidiário no processo de entendimento da condição do paciente e da implementação do tratamento, se é boa a relação médico-paciente. Sweeney e colegas (1998) assinalam que a tomada de decisões durante uma consulta ocorre em um momento naquilo que é por vezes uma doença de longo prazo e usualmente uma relação de longo prazo entre o médico e o paciente.

A aplicação de resultados de ensaios clínicos na prática clínica deve considerar esta complexidade. Decisões clínicas envolvem pessoas e, particularmente no atendimento primário, compreendê-las em seu contexto social é tão importante quanto compreender a eficácia provável da intervenção. Como diz van Weel (1996), “a qualidade do atendimento não é determinada tanto por uma intervenção ou tecnologia, por mais importantes que sejam, mas sim por sua aplicação apropriada ao paciente”. No entanto, não há vantagem em privilegiar o julgamento clínico livre e mal definido. Sackett (1996a) está correto em insistir que a tomada de decisão clínica deve envolver “o uso consciencioso, explícito e judicioso da melhor evidência disponível”. A questão importante é como, e se, podemos melhorar a qualidade das evidências para que possam ser melhor aplicadas a um dado paciente.

Dowie (1996) afirma que um sistema de apoio à decisão formal capaz de refletir a complexidade da tomada de decisão e encorajar o reconhecimento explícito dos valores diferentes que cada paciente vincula a diferentes resultados, ajudaria a melhorar o uso da evidência obtida pelos ensaios clínicos, e ele pode estar certo. No entanto, a essência do argumento para mim é que os ensaios clínicos aleatórios dizem respeito basicamente a intervenções médicas e não a pacientes. Neles, os pacientes são randomizados para permitir uma comparação da eficácia da intervenção, sem que a individualidade dos pacientes enviesse os resultados. Esta abordagem metodológica oferece para a sociedade uma proteção poderosa contra o curandeiro, e nos ajuda a eliminar ineficiências no fornecimento de cuidado médico descrito por Cochrane. Mas a minimização metodológica da informação sobre a eficiência em relação ao paciente individual deixa uma lacuna na evidência disponível para os clínicos. Esta lacuna de evidência não é nem uma nova questão, nem está restrita ao atendimento primário, mas requer uma mudança de paradigma na pesquisa e na prática clínicas. A concepção futura dos experimentos aleatórios, e o equilíbrio de investimento dos recursos escassos para pesquisa entre estudos observacionais e experimentais, devem ser guiados pela necessidade de oferecer evidência tanto sobre a aplicabilidade individual quanto a eficiência grupal. Este é um apelo para que a comunidade de pesquisa avalie métodos observacionais e veja a aplicabilidade como uma questão que é tão importante para a concepção do experimento quanto o poder estatístico. É também um apelo para os clínicos gerais usarem suas habilidades de

especialistas na avaliação dos problemas apresentados (van Weel, 1996a). Os clínicos devem aceitar o desafio de trabalhar com pacientes para melhorar o diagnóstico e o processo de tomada de decisão, de modo a prognosticar melhor o risco basal e o benefício relativo para o paciente individual em uma situação clínica complexa. Como Asher afirmou em 1961, a questão para a qual os clínicos devem encontrar uma resposta não é “isto vai funcionar?”, mas “funciona?”, de modo que em 1998 o clínico geral não deve perguntar “isso funciona para a maioria dos pacientes?”, mas sim “isto funciona para este paciente?”.

Agradeço a Frank Sullivan, Ann-Louise Kinmonth, Russel Hamilton e Chris van Weel pelos comentários úteis nas primeiras versões deste trabalho.

Referências bibliográficas

- ASHER, R. Apriority – thoughts on treatment. *Lancet* 1961, ii: 1403-04.
- BELL, J. The new genetics in clinical practice. *BMJ* 1998, 316: 618-20.
- BRADLEY, F.; WILES, R.; KINMONTH, A.L.; MANT, D.; GANTLEY, M. Development and evaluation of complex interventions in health services research. *BMJ* (in press).
- BREWEN, C.; BRADLEY, C. Patient preferences and randomised clinical trials. *BMJ* 1989, 299: 313-15.
- BURKE, P.; BAIN, J.; ROBINSON, D.; DUNLEAVY, J. Acute red ear in children: controlled trial of non-antibiotic treatment in general practice. *BMJ* 1997, 303: 558-69.
- CHARLTON, B. The scope and nature of epidemiology. *J Clin Epid* 1996, 49: 623-26.
- DEL MAR, C.; GLASZIOU, P.; HAYEM, M. Are antibiotics indicated as initial treatment for children with acute otitis media? A meta-analysis. *BMJ* 1997, 314: 1526-29.
- DOWIE, J. Evidence based, cost-effective and preference driven medicine: decision analysis based decision making is the prerequisite. *J Health Service Res Policy* 1996, 1: 104-12.
- FAHEY, T. Applying the results of clinical trials to patients in general practice: perceived problems, strengths, assumptions and challenges for the future. *Br J Gen Pract* 1998, 48: 1173-78.
- FIHN, S.; CALLAHAN, C.; MARTIN, D.; MCDONNELL, M.; HENSKOFF, J.; WHITE, R. Risk and severity of bleeding complications in elderly patients treated with warfarin. *Ann Intern Med* 1996, 124: 970-79.
- FROOM, J.; CULPEPPER, C.; GROB, P.; BARTELD, A.; BOWERS, P.; BRIDGES WEBB, C. Diagnosis of antibiotic treatment of acute otitis media. *BMJ* 1990, 300: 582-86.
- GLASZIOU, P. & IRWIG, L. An evidence based approach to individualizing treatment. *BMJ* 1995, 311: 1356-1359.
- GUYATT, G.; JUNIPER, E.; WALTER, S.; GRIFFITH, L.; GOLDSTEIN, R. Interpreting treatment effects in randomised trials. *BMJ* 1998, 316: 690-93.

- HART, J.T. What evidence do we need for evidence based medicine? *J Edip Comm Health* 1997, 51: 623-629.
- KERRIDGE, I.; LOWE, M.; HENRY, D. Ethics and evidence based medicine. *BMJ* 1998, 316: 1151-53.
- LILFORD, R. & JACKSON, J. Equipoise and the ethics of randomisation. *J R Soc Med* 1995, 88: 552-59.
- MCPHERSON, K.; BRITTON, A.; WENBERG, J. Are randomised controlled trials controlled? Patient preferences and unblind trials. *J R Soc Med* 1997, 90: 652-56.
- NAYLOR, C.D. Grey zones of clinical practice: some limits to evidence-based medicine. *Lancet* 1995, 345: 840-42.
- O'BOYLE, C. Diseases with passion. *Lancet* 1993, 342: 1126-27.
- OXMAN, A. & GUYATT, G. A consumers' guide to sub-group analysis. *Ann Intern Med* 1992; 116: 78-84.
- ROTHWELL, P.M. Can results of clinical trials be applied to all patients. *Lancet* 1995, 345: 1616-19.
- RUCKER, G. A two-stage trial design for testing treatment, self-selection and treatment preference effects. *Stats Med* 1989, 8: 477-85.
- SACKETT, D.; RICHARDSON, W.; ROSENBERG, W.; HAYNES, R.B. *Evidence based medicine: how to practice and teach EBM*. London: Churchill Livingstone, 1996.
- SACKETT, D.; ROSENBERG, W.; GRAY, J.A.M. Evidence based medicine: what it is and what it isn't. *BMJ* 1996, 312: 71-72.
- SENN, S. Random results from randomised trials. *BMJ* 1998, 317: 537.
- SILAGY, C.A. Developing a register of randomised controlled trials in primary care. *BMJ* 1993, 306: 887-900.
- SULLIVAN, F.M.; MACNAUGHTON, R.J. Evidence in consultations: intrepreted and individualised. *Lancet* 1996, 348: 941-43.
- SWEENEY, K.; GRAY, D.P.; STEELE, R.; EVANS, P. Use of warfarin in non-rheumatic atrial fibrillation. *Br J Gen Pract* 1995, 45: 153-58.
- SWEENEY, K.G.; MACAULEY, D.; PEREIRA GRAY, D. Personal significance: the third dimension. *Lancet* 1998, 35: 134-136.
- VAN WEEL, C. Chronic disease in general practice: the longitudinal dimension. *Eur J Gen Pract* 1996, 2:17-21.
- _____. Primary care: political favourite or scientific discipline? *Lancet* 1996, 348: 1431-32.