

Serviços de saúde: epidemiologia, pesquisa e avaliação

Epidemiology, research,
and health services evaluation

Iná S. dos Santos ¹
Cesar G. Victora ¹

Abstract

Health services epidemiological research generally involves testing the efficacy of new programs or strategies and the evaluation of health services quality. Repeating the evidence-based medicine model that has flourished in the clinical field, evaluations of public health programs' effectiveness has preferentially proposed and employed randomized studies as a way of guaranteeing a scientific standard of credibility in the results. A new trend among researchers of epidemiology applied to health services has challenged this premise and proposes that randomized studies be adapted to encompass the broad causal chain linking the implementation of programs to a given impact indicator. Others designs are proposed at different levels of causal inference, but sufficiently rigorous for their results to be reliable.

Services Evaluation; Health Services; Research

¹ Faculdade de Medicina,
Universidade Federal de
Pelotas, Pelotas, Brasil.

Correspondência

I. S. Santos
Faculdade de Medicina,
Universidade Federal de
Pelotas. C. P. 464, Pelotas, RS
96001-970, Brasil.
inasantos@uol.com.br

A qualidade da atenção à saúde oferecida às populações é uma questão de natureza ética. Não é ético oferecer cuidados cujo impacto não tenha sido cientificamente comprovado, ou cuja qualidade seja inferior aos padrões recomendados. Nesse sentido, abrem-se, nos serviços de saúde, duas áreas de pesquisa com naturezas complementares: investigações que visam testar a eficácia de novos procedimentos ou estratégias de cuidado e investigações voltadas para avaliar o resultado dessas intervenções sobre a saúde dos usuários e da população como um todo. Elementos fundamentais da pesquisa em serviços incluem avaliações da oferta de cuidados, expressa em termos de quantidade e qualidade de recursos de estrutura, humanos e tecnológicos; avalia-se também a utilização e a cobertura da oferta, além do impacto desta sobre a saúde das populações.

Em termos de pesquisa clínica, o movimento da Medicina Baseada em Evidências (MBE) ¹ popularizou o emprego de ensaios clínicos randomizados para testar a eficácia de novas intervenções, como drogas, vacinas ou procedimentos cirúrgicos. Tipicamente, os resultados de vários ensaios randomizados são sumarizados estatisticamente através de meta-análises (The Cochrane Collaboration: <http://www.cochrane.org>). Sem dúvida, a MBE está contribuindo substancialmente para melhorar a qualidade do atendimento clínico individual.

No entanto, esse modelo importado da pesquisa clínica vem cedendo lugar, dentro da saúde pública, a outras concepções que se apoiam, essencialmente, nas longas cadeias causais que caracterizam os fenômenos coletivos. Na pesquisa clínica, a exposição a uma droga atua específica e diretamente sobre a presença de um sintoma ou de um desfecho biológico definido, sendo curta a cadeia causal entre a exposição e o desfecho. Por outro lado, na saúde pública, o efeito de um programa somente será verificável na presença de uma série de outros eventos decorrentes da intervenção, mas modificáveis por fenômenos externos a ela (características da população, do sistema de saúde ou do ambiente). As cadeias causais em estudo, portanto, são bem mais longas e complexas. Para que o efeito de um programa de suplementação alimentar destinado a crianças desnutridas possa ser verificado, por exemplo, é necessário que o suplemento forneça os macro e micronutrientes necessários na quantidade adequada. Isoladamente, porém, essa característica do programa não será suficiente para causar um impacto, sendo necessário demonstrar que a criança-alvo foi atingida, com a regularidade necessária, e que a mãe aderiu à utilização do suplemento na forma recomendada. Logo, o efeito observável sobre o crescimento não poderá ser atribuído somente à oferta do suplemento.

Há uma corrente crescente entre intelectuais na área da epidemiologia de serviços que questiona desde a factibilidade até a necessidade da utilização de estudos randomizados, quando o objetivo é testar o impacto, em larga escala, de programas de saúde. O entendimento de que os estudos randomizados não devam ser considerados como suficientes ou exclusivos para se estabelecer o impacto de programas de saúde traz uma nova perspectiva à pesquisa na área de serviços. Nem sempre é possível, nessa área, conduzirem-se estudos randomizados. As limitações não se restringem apenas a questões éticas de se randomizarem indivíduos ou grupos de indivíduos e privá-los de um potencial benefício ou submetê-los a práticas “placebo”, mas também ao fato de, na concepção de Black ², esse desenho de estudo nem sempre ser possível, necessário, apropriado ou adequado.

Os estudos randomizados poderão ser impossíveis de conduzir em situações em que a cadeia causal ligando a intervenção ao indicador de impacto seja longa e complexa. Nesse caso, a modificação de efeito por parte de várias características externas à intervenção é altamente provável, tornando-se proibitivo o tama-

nho de amostra necessário para dar conta de todos esses aspectos. Um corolário dessa observação é que o uso de técnicas de meta-análise – usualmente apropriadas para resumir resultados de estudos clínicos – pode ser pouco recomendável para a avaliação de programas complexos. Nesses últimos, o efeito de fatores contextuais locais é tão importante, que a premissa de homogeneidade de resultados entre diferentes estudos deixa de ser válida.

Os estudos randomizados poderão ainda ser desnecessários. Há inúmeros exemplos na área clínica de tratamentos cujo impacto sobre a doença foi tão dramático, que dispensaram a necessidade de estudos controlados para que sua eficácia fosse demonstrada. Foi o caso da penicilina no tratamento das pneumonias, do digital no tratamento da insuficiência cardíaca e da insulina no tratamento do diabetes insulino-dependente. Na saúde pública, a redução do número de óbitos de motociclistas envolvidos em acidentes de trânsito, após a introdução da lei que obrigou o uso de capacetes entre usuários de motocicletas, é um exemplo que dispensa a realização de estudos randomizados para que seu impacto seja aceito. O próprio John Snow, ao remover a bomba da fonte de água contaminada em Broad Street, realizou um estudo do tipo antes-e-depois, sem grupo-controle, cujo efeito foi tão marcante, que o transformaram no fundador da epidemiologia.

Nem sempre é apropriado conduzirem-se estudos randomizados. Os estudos de fase IV de medicamentos são um exemplo. Pode ser tão baixa a incidência de efeitos colaterais de drogas cuja eficácia foi testada em grandes ensaios randomizados, que a monitorização de tais efeitos depende de estudos observacionais em larga escala.

As limitações de validade interna e externa dos ensaios randomizados acentuam-se quando o objeto de estudo situa-se no campo da saúde pública, podendo torná-los inadequados. Abordamos os problemas de validade interna em outra publicação ³, mas a principal limitação diz respeito à validade externa. Estudos randomizados são normalmente conduzidos sob condições ideais, em que tanto os provedores de serviço, quanto os recipientes da intervenção, recebem uma série de medidas de apoio para que sua aderência ao programa seja maximizada. Por exemplo, os provedores são intensamente treinados, de acordo com protocolos padronizados; o suprimento de insumos é plenamente adequado e as condições de trabalho são ideais. Já os recipientes são freqüentemente contemplados com programas educacionais e com medidas de apoio, como forneci-

mento gratuito de medicações e exames complementares, auxílio financeiro direto ou indireto (por exemplo, auxílio para deslocamento até o serviço de saúde ou visitas domiciliares), a fim de maximizar sua adesão. Como a dose de intervenção assim recebida pelos usuários é bastante superior à recebida sob condições de rotina, os estudos randomizados testam normalmente a eficácia e não a real efetividade do programa. Mesmo estudos randomizados delineados para testar efetividade não conseguem alcançar plenamente seus objetivos, pois a mera presença de uma equipe de coleta de dados e o conhecimento por parte dos provedores de que estão sendo avaliados dentro de um estudo contribuem para o efeito Hawthorne ⁴, em que os resultados obtidos são normalmente superiores aos observados na vida real.

Pelos motivos expostos acima, os estudos randomizados nem sempre poderão avaliar programa de saúde pública, necessitando ser complementados ou substituídos por outros delineamentos.

Uma das principais contribuições teóricas recentes no campo da epidemiologia aplicada aos serviços de saúde foi dada por Habicht et al. ⁵, em seu artigo *Evaluation Designs for Adequacy, Plausibility and Probability of Public Health Programme Performance and Impact*, no qual são definidos dois eixos que norteiam a avaliação de programas de saúde e nutrição. O primeiro eixo refere-se ao tipo de indicador a ser utilizado, o que depende, essencialmente, do que está sendo medido: se o processo, o resultado ou ambos os aspectos do programa. Indicadores de oferta, utilização e cobertura analisam o processo; indicadores de impacto, o resultado. Na avaliação de um programa de pré-natal, por exemplo, a oferta poderá ser medida através de indicadores que avaliem os equipamentos, a área física, os recursos humanos e o cumprimento das normas do programa durante as consultas. O número de consultas pré-natais e a idade gestacional de início do acompanhamento pré-natal serão indicadores de utilização. A cobertura do programa será medida da maneira convencional – proporção de gestantes da área de abrangência do programa sendo atendidas por este. O impacto poderá ser medido utilizando-se vários indicadores, como tipo de parto, peso ao nascer e duração da amamentação.

A decisão sobre que indicadores devem ser medidos dependerá de quem utilizará os resultados e das decisões a serem tomadas em decorrência da avaliação. Assim, na avaliação de um programa de suplementação alimentar na infância, por exemplo, a chefia de um posto de

saúde precisará saber apenas se o estoque disponível do suplemento será suficiente para atender a demanda da comunidade (indicadores de oferta). Para a Secretaria Municipal de Saúde, poderá ser importante saber se a maioria das crianças do município recebe o suplemento regularmente (indicadores de utilização). Para a Secretaria Estadual de Saúde, o mais importante poderá ser identificar as características do programa nos municípios que logram maiores coberturas (indicador de cobertura). E, finalmente, ao Ministério da Saúde interessa se o investimento está reduzindo a mortalidade entre os menores de cinco anos de idade (indicadores de impacto).

O segundo eixo refere-se ao tipo de inferência necessária para se afirmar que os resultados observados, tanto de processo quanto de impacto, foram efetivamente decorrentes da exposição à intervenção. Habicht et al. ⁵ identificam avaliações em três níveis de inferência causal: adequação, plausibilidade e probabilidade. Avaliações de adequação investigam se os objetivos do programa foram alcançados. Avaliações de plausibilidade, se os objetivos foram alcançados devido ao programa. E avaliações de probabilidade estimam, como o próprio nome indica, qual a probabilidade estatística de que o programa tenha realmente efeito.

Avaliações de adequação são mais simples e devem ser feitas sempre antes de estudos mais complexos e, conseqüentemente, dispendiosos. Avaliam se os objetivos foram alcançados antes de atribuí-los ao programa. Se uma intervenção é sabidamente eficaz, então uma avaliação de adequação da cobertura poderá ser suficiente.

Avaliações de plausibilidade analisam se o efeito observado deveu-se ao programa ou a fatores externos, exigindo, portanto, algum tipo de grupo-controle (interno, histórico ou externo), sem utilizar, no entanto, randomização. Grande parte dos estudos brasileiros na área de avaliação de serviços apresenta esse tipo de delineamento. Em um estudo para avaliar o impacto do Programa do Leite em Alagoas ⁶, por exemplo, uma avaliação de plausibilidade comparou crianças desnutridas que recebiam o suplemento, as quais residiam em dez municípios com alta cobertura do programa, com crianças desnutridas não atingidas pelo programa, residentes em outros dez municípios com baixa cobertura. Nesse tipo de avaliação, independente do tipo de grupo-controle, diferenças externas ao programa precisam ser descartadas. A renda familiar e escolaridade do pai, por exemplo, eram menores entre as crianças atendidas pelo programa. Essa é uma caracte-

rística positiva do ponto de vista de adequação do foco do programa. Contudo, se essas variáveis se mostrarem associadas ao desfecho, precisariam ser controladas na análise, para que o impacto real do programa pudesse ser estimado.

Avaliações de probabilidade, importantes na fase de desenvolvimento de novas intervenções, requerem alocação aleatória das unidades de estudo para que estas recebam ou não o programa. A eficácia do aconselhamento nutricional para crianças de 0 a 18 meses, por exemplo, foi testada em uma avaliação de probabilidade ⁷. Quatorze postos de saúde foram escolhidos aleatoriamente e seus médicos, treinados em aconselhamento nutricional (grupo intervenção); os médicos dos outros 14 postos de saúde do município constituíram o grupo-controle. Acompanharam-se as crianças atendidas pelos médicos de ambos os grupos e avaliou-se o impacto do treinamento. Como mencionado anteriormente, na área da saúde pública, as avaliações de probabilidade não dispensam análises de adequação e plausibilidade. No estudo citado, foi necessário demonstrar que o maior conhecimento, habilidade e desempenho dos médicos treinados conduziu a um número suficiente de mães recebendo o aconselhamento nutricional, as quais modificaram seu comportamento em relação à alimentação infantil, o que resultou em melhora da dieta e conseqüente melhora no estado nutricional da criança.

Uma vez demonstrada a eficácia da intervenção, no entanto, as avaliações subseqüentes do programa podem usar desenhos de plausibilidade e adequação. Por exemplo, se a eficácia de uma vacina foi adequadamente comprovada em estudos de probabilidade, é suficiente documentar apenas as tendências temporais na cobertura e na incidência da doença em questão. O controle do sarampo no Brasil é um exemplo claro de uma avaliação de adequação que não deixa dúvidas sobre o impacto da vacina (<http://www.datasus.gov.br>).

Além disso, muitas decisões acertadas podem ser tomadas com base em análises de adequação e de plausibilidade. Avaliações mais

complexas, por exemplo sobre impacto de um programa, somente devem ser feitas se estudos mais simples tiverem mostrado que o programa está sendo adequadamente ofertado e que a cobertura populacional necessária foi atingida.

Que delineamentos utilizar na avaliação de programas ou serviços de saúde? Os delineamentos ideais dependem da natureza dos programas e da precisão das estimativas a serem obtidas. Avaliações de adequação podem ser realizadas através de estudos transversais (por exemplo, para avaliar cobertura ou qualidade de serviços) ou longitudinais (por exemplo, para avaliar tendências nos indicadores de impacto). Avaliações de plausibilidade poderão ser feitas através de delineamentos transversais ou longitudinais controlados; nas avaliações de impacto, o delineamento caso-controle poderá ser também empregado. Avaliações de probabilidade só admitem estudos longitudinais controlados, para análise tanto de processo, quanto de impacto.

Em resumo, os estudos randomizados seguem sendo importantes na determinação da eficácia de novas intervenções. O que se advoga, de novo, é que seus resultados precisam ser complementados, na área da saúde pública, por estudos observacionais que testem a efetividade das intervenções sob condições de rotina. Os resultados de estudos não randomizados precisam ser alvo de tanta atenção quanto a que tem sido dedicada, nos últimos anos, aos estudos randomizados. Uma iniciativa recente irá contribuir para esse fim. Assim como os estudos randomizados devem ser redigidos de acordo com uma série de normas editoriais – estabelecidas pelo CONSORT, ou *Consolidated Standards of Reporting Trials* (<http://www.consort-statement.org>) –, os estudos do tipo plausibilidade agora também dispõem de uma série de normas similares – a declaração TREND ⁸. Espera-se que essa iniciativa contribua para que melhore a qualidade da divulgação de estudos não randomizados e para que seus resultados forneçam subsídios para políticas de saúde baseadas em evidências.

Resumo

A pesquisa epidemiológica na área de serviços de saúde envolve geralmente a experimentação da eficácia de novos programas ou estratégias e a avaliação da qualidade dos serviços de saúde. Repetindo o modelo da medicina baseada em evidência, que floresceu na área clínica, as avaliações de efetividade de programas de saúde pública têm proposto e empregado, preferentemente, estudos randomizados, como forma de garantir um padrão científico de credibilidade nos resultados. Uma nova tendência entre os intelectuais da epidemiologia aplicada aos serviços de saúde questiona essa premissa e propõe que os estudos randomizados sejam adaptados, para compreender a ampla cadeia causal que vincula a implementação de um programa a um determinado indicador de impacto. Outros desenhos, em níveis diferentes de inferência causal, mas suficientemente rigorosos para que seus resultados sejam confiáveis, são também propostos.

Avaliação dos Serviços; Serviços de Saúde; Pesquisa

Colaboradores

I. S. Santos estruturou o conteúdo do texto e redigiu a primeira versão do manuscrito. C. G. Victora revisou e implementou modificações à versão final do manuscrito.

Referências

1. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence-based medicine: what it is and what it isn't. *BMJ* 1996; 312:71-2.
2. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996; 312:1215-8.
3. Victora CG, Habicht JP, Bryce J. Evidence-based public health: moving beyond randomized trials. *Am J Public Health* 2004; 94:400-5.
4. Fletcher RH, Fletcher SW, Wagner EH. *Epidemiologia clínica: bases científicas da conduta médica*. Porto Alegre: Editora Artes Médicas; 1996.
5. Habicht JP, Victora CG, Vaughan JP. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *Int J Epidemiol* 1999; 28:10-8.
6. Gigante DP, Santos IS, Coitinho D, Valle NJ, Haisma H, Valente GCM. Avaliação do impacto do Programa Nacional do Leite em Alagoas, através de métodos isotópicos: aspectos metodológicos e resultados preliminares. *Rev Bras Epidemiol* 2002; 5:63-70.
7. Santos I, Victora CG, Martines J, Gonçalves H, Gigante DP, Valle NJ, et al. Nutrition counseling increases weight gain among Brazilian children. *J Nutr* 2001; 131:2866-73.
8. Des Jarlais DC, Lyles C, Crepaz N, The Trend Group. Improving the reporting quality of non-randomized evaluations of behavioral and Public Health interventions: the TREND statement. *Am J Public Health* 2004; 94:406-15.

Recebido em 15/Mar/2004

Aprovado em 18/Mai/2004