

Avaliação

conceitos e métodos

Astrid Brousselle | François Champagne
André-Pierre Contandriopoulos | Zulmira Hartz
Organizadores

EDITORA



HOC RUIZ

Copyright © 2011 dos autores
Todos os direitos desta edição reservados à
FUNDAÇÃO OSWALDO CRUZ / EDITORA

Tradução de *L'Évaluation: concepts et méthodes*, publicado por
Les Presses de l'Université de Montréal em 2009. Reprodução autorizada.

Capa, projeto gráfico
Carlota Rios

Imagem da capa
A partir da pintura 'Bloom' de Paul Klee, de 1934.

Revisão
Irene Ernest Dias

Catálogo na fonte
Instituto de Comunicação e Informação Científica e Tecnológica em Saúde
Biblioteca de Saúde Pública

B876a Brousselle, Astrid (Org.)
Avaliação: conceitos e métodos. / organizado por Astrid
Brousselle... [et al]; tradução de Michel Colin. - Rio de Janeiro :
Editora Fiocruz, 2011.
292 p., il.

ISBN: 978-85-7541-218-3

1. Avaliação em Saúde - métodos. 2. Avaliação de Programas
e Projetos de Saúde. 3. Planejamento Estratégico. I. Champagne,
François (Org.). II. Contandriopoulos, André-Pierre. III. Hartz,
Zulmira. IV. Título.

CDD - 22.ed. - 362.1

2011
EDITORA FIOCRUZ
Av. Brasil, 4036 - Térreo - sala 112 - Manguinhos
21040-361 - Rio de Janeiro - RJ
Tels: (21) 3882-9039 e 3882-9041
Telefax: (21) 3882-9006
e-mail: editora@fiocruz.br
www.fiocruz.br/editora

8

A Análise dos Efeitos

*François Champagne, Astrid Brousselle,
André-Pierre Contandriopoulos e Zulmira Hartz*

A análise dos efeitos examina as relações causais entre uma intervenção e seus efeitos. Trata-se, provavelmente, da avaliação mais clássica, e continua a ser amplamente utilizada.

A análise dos efeitos faz parte da pesquisa avaliativa. Duas diferenças fundamentais a distinguem da apreciação normativa dos resultados. Primeiro, a análise dos efeitos se interessa pelo conjunto dos efeitos possíveis de uma intervenção, enquanto a apreciação normativa se interessa apenas pelos efeitos almejados pela intervenção. Aliás, o presente capítulo mostrará como os efeitos de uma intervenção são, frequentemente, mais amplos que os objetivos almejados por seus idealizadores. Segundo, a análise dos efeitos tem por objeto a relação de causalidade entre os serviços e os efeitos, enquanto a apreciação normativa não examina essa relação e se limita a comparar os efeitos observados com os resultados esperados.

Duas grandes perguntas guiam a análise dos efeitos: “Que efeitos medir?” e “Os efeitos observados resultam realmente da intervenção?”.

Que Efeitos Medir?

Qualquer intervenção produz efeitos, quer sejam positivos, negativos ou nulos – pois deve-se dizer que até mesmo a ausência de mudança constitui

um efeito. Alguns efeitos são desejados e outros, não; alguns são previsíveis, outros não o são.

Os diferentes tipos de efeitos

Vimos que a análise lógica da intervenção permite identificar os efeitos desejados e previsíveis. Graças à revisão da literatura e à consulta aos especialistas e aos agentes da intervenção, o modelo lógico permite ir além da cadeia causal principal e identificar grande parte dos efeitos não desejados, mas previsíveis. Em compensação, o modelo lógico teórico nem sempre consegue revelar efeitos não previsíveis. O exemplo a seguir permite compreender melhor o que é um efeito não previsível e não desejado. Trata-se de uma intervenção cujo objetivo era promover o uso do capacete pelos ciclistas. O efeito desejado e previsível era a redução do número de óbitos e traumatismos cranianos imputáveis aos acidentes. Ora, observou-se que os motoristas adotavam uma conduta mais ousada à vista de um ciclista que usava capacete do que quando passavam por um ciclista sem capacete. Portanto, havia um risco de que a intervenção contribuísse para o aumento do número de acidentes, o que constituía, evidentemente, um efeito não desejado e não previsível.

De fato, pode acontecer de uma intervenção provocar efeitos perversos que neutralizem sobremaneira os efeitos positivos. Nesse caso, o fato de medir unicamente os efeitos positivos não permite obter uma imagem válida dos efeitos. Estes últimos se estendem muito além dos objetivos, e é importante ficar atento ao conjunto dos efeitos possíveis, sejam eles desejados ou não, previsíveis ou não. Para identificar os efeitos não previsíveis de uma intervenção, é preciso levar em conta a experiência dos atores e dos participantes, e não apenas os textos especializados, pois nestes últimos os efeitos nulos ou indesejáveis são sub-representados em razão de um viés de publicação.

Vimos no capítulo 3 que uma intervenção é uma longa cadeia causal de ações que liga estruturas, processos e resultados. A questão que se coloca é a seguinte: onde medir os efeitos dessa longa cadeia? É impossível para um avaliador medir todos os aspectos de uma cadeia de ação. Consequentemente, ele deve estabelecer os que considera efeitos e, entre estes, aqueles que considera oportuno medir. Retomando o exemplo da intervenção sobre o uso do capacete pelos ciclistas, os seus objetivos de produção eram modificar um dos fatores que predisõem aos acidentes fatais para, aumentando-se o uso do capacete,

afinal, reduzir os óbitos e traumatismos cranianos. Neste exemplo, que efeitos devem ser medidos? E em qual momento se deve medir?

Nem sempre é preciso ir até o fim da cadeia causal para que se possa medir efeitos. Tudo depende da robustez das hipóteses subjacentes descritas no modelo lógico. Por exemplo, quando da avaliação de um programa de informação sobre o tabaco e suas consequências, destinado a reduzir a incidência do câncer do pulmão, não é necessário mostrar que as pessoas que estiveram expostas à intervenção têm menos câncer. A relação entre o tabagismo e a incidência do câncer do pulmão está suficientemente bem estabelecida para que seja suficiente medir a redução do consumo de tabaco. No entanto, é possível medir a modificação dos conhecimentos sobre o tabaco e suas consequências, que constitui um objetivo anterior àquele da redução do tabagismo. Todavia, neste último caso, como o vínculo causal entre as modificações dos conhecimentos e os comportamentos de consumo do tabaco não é robusta, não se poderá concluir por uma intervenção eficaz sem ir até pelo menos a medida do comportamento de tabagismo. Assim, é possível escolher analisar diferentes efeitos na cadeia causal, mas não se deve esquecer que nem todos têm o mesmo grau de pertinência. Com o auxílio do modelo lógico, é possível considerar os efeitos pertinentes (a serem medidos) partindo-se daqueles que têm um valor intrínseco, para ir até aqueles cujo impacto foi bem demonstrado.

É muito importante também escolher adequadamente o momento em que se medem os efeitos de uma intervenção. Por exemplo, observa-se que os programas de informação destinados a reduzir o tabagismo ocasionam mudanças de comportamentos logo a partir do fim da intervenção. Mas observa-se também que essas mudanças de comportamento, isto é, os efeitos, não são muito duradouras. Além disso, os efeitos a curto prazo não são necessariamente os mesmos que os efeitos a longo prazo. O tempo e determinados acontecimentos podem provocar uma ampliação ou, ao contrário, uma atenuação dos efeitos. Por outro lado, quanto mais longo for o lapso de tempo entre o fim de uma intervenção e sua avaliação, mais difícil se tornará associar os efeitos com a intervenção. Levando-se em conta o fato de que os efeitos não são estáveis no decorrer do tempo, torna-se necessário se perguntar também qual é o momento mais propício para medi-los.

Alguns autores não falam de efeitos, mas de resultados primários e resultados secundários. Estabelecem, assim, uma distinção entre a cadeia causal principal e outros objetivos menos importantes. Os economistas distinguem,

por sua vez, as externalidades – isto é, os efeitos que atingem uma pessoa ou um grupo não diretamente alvo da intervenção – dos efeitos principais. Pouco importa a terminologia privilegiada, todos esses elementos deveriam poder ser integrados no modelo lógico teórico.

Finalmente, uma advertência se impõe. É preciso poder sempre justificar, no plano teórico, a escolha dos efeitos que se quer medir. É um erro frequente e grave escolher os efeitos que se quer medir unicamente em função da facilidade de sua medida – entre outras coisas porque existem instrumentos validados –, mais que em função do que é conceitualmente desejável medir.

Os tipos de eficácia

Os efeitos de uma intervenção podem ser medidos em laboratório ou diretamente na população que é seu objeto de atuação. O contexto em que se escolhe medir os efeitos influi necessariamente sobre os resultados obtidos (Quadro 1). Assim, a eficácia de uma intervenção poderá parecer diferente conforme seus efeitos forem medidos em laboratório ou diretamente entre a população.

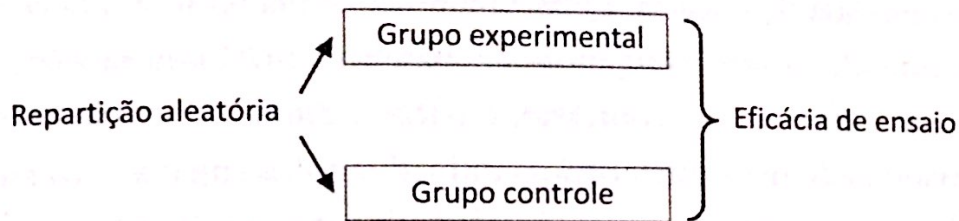
Quadro 1 – Os tipos de eficácia

CONTEXTO DA PESQUISA	EFICÁCIA MEDIDA	ORIGEM DAS DIFERENÇAS NAS EFICÁCIAS MEDIDAS
Laboratório	Eficácia teórica	Variabilidades biológicas interindividuais
Ensaios clínicos	Eficácia de ensaio	Diferenças na seleção dos pacientes (utilizadores) e nos comportamentos (observância) dos pacientes e dos profissionais
Prática normal	Eficácia de utilização	Diferenças na acessibilidade da intervenção
Comunidade, população	Eficácia populacional	

Em laboratório, as condições de experimentação são controladas para que todos os parâmetros que possam exercer alguma influência sobre os efeitos, à exceção da intervenção estudada, sejam levados em conta. O dispositivo experimental permite, sobretudo, eliminar a influência das particularidades e dos comportamentos dos sujeitos. A eficácia observada em tais condições é, entretanto, uma eficácia teórica.

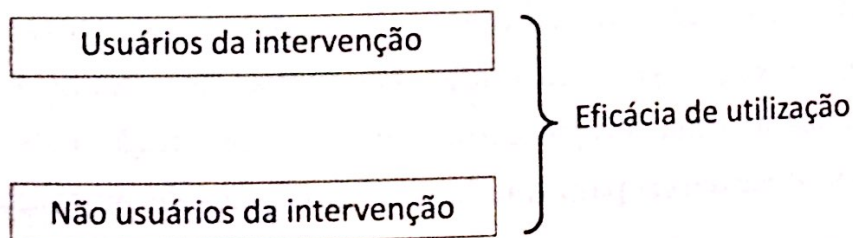
Os ensaios clínicos constituem um segundo contexto de experimentação. Nesse tipo de pesquisas, o grupo experimental e o grupo controle são selecionados de modo aleatório para se garantir a similitude dos dois grupos, e as condições de administração da intervenção são assim controladas. Os únicos elementos que não são levadas em conta são as variações biológicas entre os indivíduos. Nessas condições, a eficácia medida é chamada eficácia de ensaio.

Figura 1 - A eficácia de ensaio



Pode-se medir também a eficácia de uma intervenção em um contexto de prática normal. Nesse caso, os utilizadores não são selecionados em função de características individuais preestabelecidas (condições de inclusão e de exclusão dos sujeitos nos ensaios clínicos). As condições de administração da experimentação não são mais controladas, e podem ocorrer variações importantes de comportamento desses usuários (observância do tratamento), bem como variações de comportamento dos profissionais. A eficácia, então, é dita de utilização.

Figura 2 - A eficácia de utilização



Finalmente, a eficácia pode ser medida na escala da comunidade, da população. Falamos, então, de eficácia populacional. Aqui, além das variações de comportamento dos usuários e dos profissionais, pode haver também diferenças na acessibilidade ao tratamento, o que pode reduzir consideravelmente a eficácia da intervenção. É o caso, por exemplo, dos cursos pré-natais, que

têm excelente eficácia de utilização. As mulheres que seguem esses cursos apresentam maior satisfação com os serviços, e seus bebês têm peso mais elevado ao nascer. Todavia, a eficácia populacional desse tipo de programa é muito baixa, e isso simplesmente porque não se consegue alcançar as mulheres que mais se beneficiariam dele. A eficácia calculada com base nos efeitos na população-alvo é, portanto, mais baixa que a eficácia medida unicamente entre as usuárias.

Em resumo, o contexto de medida dos efeitos tem influência determinante sobre a medida da eficácia. Quanto mais controladas forem as condições, mais selecionados forem os sujeitos com base em critérios homogêneos, mais elevada será a eficácia. Ao contrário, a eficácia diminui quando se introduz certa variabilidade nas características individuais dos sujeitos, nos modos de prática dos profissionais e na acessibilidade da intervenção. O avaliador deve se perguntar sobre a pertinência de sua escolha, no que concerne ao contexto de medida dos efeitos, e a quem se dirige a avaliação dos efeitos. Quando se trata de elaborar uma nova intervenção (por exemplo, um medicamento), é preferível escolher uma eficácia de ensaio ou uma eficácia teórica. Ao contrário, quando se trata de medir os efeitos de uma intervenção bem reconhecida e os desafios têm por objeto a aplicabilidade a um contexto particular, é preferível medir a eficácia de utilização ou populacional.

Os Efeitos Observados Resultam Realmente da Intervenção?

Na análise dos efeitos, é essencial certificar-se de que os efeitos medidos estão realmente associados à intervenção. Essa exigência está diretamente ligada à questão da causalidade. Na elaboração do modelo lógico teórico, as relações causais da intervenção são valorizadas. Na análise dos efeitos, é preciso que se possa encontrar uma relação causal entre a intervenção e os efeitos. Três condições são necessárias para que haja uma relação de causalidade:

- a causa deve preceder o efeito (dimensão temporal);
- as duas variáveis devem estar ligadas empiricamente (presença de variações concomitantes);
- a relação entre as duas variáveis não pode ser explicada por uma terceira variável.

Teoricamente, uma relação de causalidade entre duas variáveis só se estabelece se a relação entre X e Y implica um determinismo absoluto, isto é, se

reúne as condições de necessidade e de suficiência. A condição de necessidade implica que X deve acontecer para que Y aconteça, enquanto a condição de suficiência traduz a ideia de que se X acontecer, então Y acontece. Em ciências sociais, o modelo probabilista predomina, o que significa que a causalidade se refere também a situações de determinismo não absoluto (suficiente ou necessário) ou a variações concomitantes. Em uma relação de determinismo não absoluto, ou X é necessário mas não suficiente para que Y aconteça, ou X não é necessário, mas suficiente para a produção de Y. Em uma relação concomitante, X não é nem necessário nem suficiente para que Y aconteça. Todavia, se X estiver presente, é mais provável que Y aconteça ou, se X estiver ausente, é mais provável que Y esteja ausente.

Para certificar-se de que há uma covariação entre dois fatores e de que esta não se deve a um terceiro fator, seria necessário poder comparar aquilo que se produziria se uma causa estivesse presente com aquilo que aconteceria se ela estivesse ausente, desde que todas as outras condições fossem mantidas. Porém, essa comparação é impossível. Por isso é preciso dar um salto inferencial que consiste em manipular experimentalmente os demais fatores – sujeito, contexto, tempo. Graças a essa comparação imperfeita, é possível aproximar-se muito da comparação ideal, mas impossível. Poder-se-ia, por exemplo, fazer e não fazer alguma coisa em dois grupos diferentes ou, então, em dois momentos diferentes.

Existem dois tipos de experimentações: a experimentação provocada e a experimentação invocada. A experimentação provocada é a manipulação ativa e intencional de uma ou de diferentes variáveis independentes – neste caso, da intervenção – para observar seus efeitos sobre outras variáveis, as variáveis dependentes. A experimentação invocada utiliza variações naturais, não organizadas pelo pesquisador, segundo a lógica do método experimental.

As Questões de Validade

Para maximizar a validade da experimentação, é preciso procurar responder a duas perguntas principais:

1. A relação observada na experiência é verdadeira? Para responder positivamente, é preciso ter certeza, por um lado, de que não existe viés (erro sistemático) que decorre das hipóteses rivais e dos fatores de confusão, que ameaçam a validade interna da experimentação, e, por outro lado,

de que não existe erro aleatório, ligado à estabilidade da relação observada. Neste último caso, o problema está situado no plano da validade da conclusão estatística.

2. Em que circunstâncias a relação é verdadeira? Componentes empíricos (população, meio) ou teóricos (conceitos, construtos) podem ameaçar a validade externa da experimentação e limitar a extrapolação dos resultados.

O pesquisador deve escolher as condições de experimentação – é o que se chama de fazer um projeto de pesquisa – para pôr à prova o modelo teórico, isto é, para testar as hipóteses. Um projeto de pesquisa determina o que será feito, a quem e quando. Ele deve ser estabelecido de modo a que a pesquisa produza os resultados cuja validade seja a maior possível.

A validade de uma avaliação é o grau de correspondência entre as conclusões de uma avaliação e o “valor real” daquilo que se avalia (Brinberg & McGrath, 1982). A validade de uma avaliação constitui, de certo modo, sua qualidade. Seja qual for o tipo de avaliação escolhido, é preciso certificar-se de que o método de pesquisa utilizado é rigoroso, de modo a produzir resultados válidos. Distinguem-se, habitualmente, quatro tipos de validade: a dos instrumentos e da estratégia de medida; a validade interna; a validade externa e, enfim, a validade da conclusão estatística.

A validade dos instrumentos e da estratégia de medida

A validade de um instrumento é sua capacidade real de medir corretamente o que ele se destina a medir. Existem três tipos de validade de medida.

A *validade de conteúdo* corresponde à capacidade de um instrumento para delimitar todas as dimensões do conceito a medir. Por exemplo, uma medida da melhoria do estado de saúde de uma população que levasse em conta apenas a taxa de mortalidade geral teria muito pouca validade de conteúdo. A validade de conteúdo da medida de uma variável pode ser otimizada decompondo-se a variável ou o conceito em tantas dimensões quanto possível. As técnicas de consenso de grupo, como o grupo nominal, podem ser utilizadas com esse objetivo.

A *validade de critério* corresponde à capacidade real de um instrumento para prever ou estimar as variações de um fenômeno. Quando um instrumento de medida é utilizado para prever um fenômeno (ou uma variável), o fenômeno que se tenta prever se torna o “critério” da validade do instrumento.

Por exemplo, um exame de admissão na universidade é válido se, e somente se, consegue prever exatamente o êxito ulterior dos estudantes admitidos na universidade. Nesse caso, em que o critério só pode ser observado e verificado ulteriormente, fala-se de validade preditiva. Quando o critério já existe - por exemplo, quando se quer utilizar os escores de quociente intelectual para estimar as diferenças no desempenho dos estudantes -, fala-se, então, em validade concomitante.

A *validade de construto* corresponde ao rigor com que um instrumento permite conceber e tornar operantes as variáveis estudadas (saúde, inteligência, acessibilidade etc.). Essas variáveis são construções teóricas nascidas da generalização ou do agrupamento de fenômenos. Essas abstrações não existem concretamente e não são diretamente observáveis. Quando se tenta medi-las, é necessário se perguntar até que ponto os instrumentos e a estratégia de medida permitem bem delimitar tais abstrações teóricas. Mede-se verdadeiramente a saúde, a acessibilidade ou a inteligência? É possível estimar a validade de construto examinando-se as correlações entre várias medidas diferentes de um mesmo conceito (abordagem *trait validation*), construindo-se uma rede de intercorrelações entre várias medidas de vários conceitos (abordagem *multitraço-multimétodos*) ou examinando-se as correlações entre vários conceitos cujas inter-relações se postulam antecipadamente (abordagem *nomológica*).

A validade interna

A validade interna corresponde à certeza de que as variações observadas nas variáveis dependentes, quando da avaliação, são verdadeiramente devidas à intervenção. Em outras palavras, a avaliação deve permitir certificar que é de fato a intervenção (variável independente) que está na origem das mudanças observadas. É preciso, para tanto, certificar-se de que o desenho de pesquisa adotado permite eliminar os diferentes vieses que possam afetar a validade interna. Tais vieses são descritos abaixo.

Quadro 2 - Vieses que podem influenciar a validade interna de uma pesquisa experimental

1. História	Viés ocasionado pelo surgimento de um fato que influencia a variável dependente durante o desenrolar do estudo.
2. Maturação	Viés ocasionado pelo envelhecimento, pelo cansaço, pela experiência adquirida pelos sujeitos entre o pré-teste e o pós-teste.
3. Familiarização com o teste	Viés que se produz quando os sujeitos aprendem a responder melhor (ou menos bem) a um questionário que lhes é submetido várias vezes durante o estudo.
4. Mortalidade experimental	Viés ocasionado por taxas de abandono diferentes no grupo experimental e no grupo controle.
5. Regressão para a média	Viés que aparece quando se escolhe uma população experimental ou controle a partir dos resultados obtidos quando do pré-teste. Aqueles que tiveram os melhores resultados têm, simplesmente em função do acaso, possibilidades de ver seu desempenho diminuir (se aproximar da média) no pós-teste, e vice-versa.
6. Seleção	Viés ocasionado pelas diferenças que possa haver entre o grupo experimental e um grupo controle não equivalente.
7. Interações	Viés que sobrevém quando interações acontecem entre a seleção e os outros vieses, sobretudo no que diz respeito à história, à maturação e à medida dos efeitos.
8. Medida dos efeitos	Viés ligado a mudanças nos instrumentos de medida durante o desenrolar da pesquisa ou à habituação dos observadores aos instrumentos de medida, isto é, a uma mudança no modo como os instrumentos de medida são utilizados entre o início e o fim do estudo.

Fontes: Campbell & Stanley (1963); Contandriopoulos *et al.* (1990).

A validade externa

A validade externa é a capacidade de generalizar para outras situações a relação demonstrada entre um programa e seus efeitos. Em outros termos, é necessário poder certificar-se de que as conclusões da avaliação são generalizáveis para outros meios, outros períodos e outras populações. O Quadro 3 apresenta os vieses que devem ser controlados para se obter uma validade externa adequada.

Quadro 3 – Vieses que podem influenciar a validade externa de uma pesquisa experimental

VIESES ASSOCIADOS À REATIVIDADE DOS SUJEITOS À SITUAÇÃO EXPERIMENTAL	
9. Contágio	Viés ligado à falta de independência entre o grupo controle e o grupo experimental.
10. Reações compensatórias	Vieses ligados às mudanças de comportamento da população do grupo testemunha após a certeza de que não receberão o tratamento. As reações compensatórias podem ser positivas (fazer tão bem ou melhor que o grupo experimental) ou negativas (desmoralização).
11. Desejo de agradar ao avaliador	Viés ocasionado pelo desejo dos sujeitos estudados de parecerem em boa saúde, de terem um comportamento sadio, de serem bem avaliados.
VIESES ASSOCIADOS A UM CONTROLE IMPERFEITO DA INTERVENÇÃO	
12. Relação causal ambígua	Viés que pode existir quando o modelo teórico da pesquisa é insuficiente.
13. Intervenções compensatórias	Viés ocasionado pelos atores que podem tentar compensar a ausência de tratamento do grupo controle por uma atenção maior ou quando, por preocupação de equidade administrativa, o grupo controle recebe diferentes formas de compensação.
14. Expectativa do experimentador	Viés ocasionado pelas expectativas do experimentador que espera este ou aquele tipo de resultado. Neste caso, não se sabe o que se deve à própria intervenção ou o que se deve às expectativas do experimentador.
15. Interação entre a intervenção e a situação experimental	Viés que ocorre na seguinte situação: com frequência, a intervenção analisada constitui um conjunto complexo que pode, no âmbito de uma pesquisa experimental, ter características especiais (entusiasmo do promotor de uma experiência piloto, favorável à intervenção, momento particularmente pouco propício para um novo modo de atuar...). Quando isso ocorre, pode ser espinhoso separar os efeitos da intervenção propriamente dita daqueles criados pelas particularidades da situação experimental.
16. Interação entre os diferentes componentes de uma intervenção	Viés associado à dificuldade para separar os efeitos dos componentes de uma intervenção. É para corrigir esta situação que são introduzidos placebos nas pesquisas experimentais. Isso permite, por exemplo, separar o efeito de um medicamento propriamente dito do efeito associado ao sentimento de se sentir cuidado (efeito placebo).
17. Interação entre as observações e a intervenção	Viés que se produz quando, em uma situação experimental, o questionário ganha mais importância que a intervenção, o que pode interagir com a própria intervenção para multiplicar ou diminuir seus efeitos, e os sujeitos estudados ficam então sensibilizados para o conteúdo da intervenção.
VIESES ASSOCIADOS À SELEÇÃO DOS SUJEITOS	
18. Interação entre a seleção dos sujeitos e a intervenção	Viés ocasionado pelo caráter não representativo dos sujeitos que são objeto do estudo. Este viés se manifesta em especial quando uma pesquisa é feita com sujeitos voluntários. Torna-se, então, delicado saber o que se deve ao voluntariado e o que se deve à própria intervenção.

A validade da conclusão estatística

A validade da conclusão estatística concerne à relação estatística entre uma causa e um efeito presumidos (Cook *et al.*, 1982). Em outros termos, será possível inferir estatisticamente que existe (ou não) uma associação entre, por um lado, variações observadas nas variáveis dependentes e, por outro, o programa avaliado? A avaliação é suficientemente precisa, sensível e potente para detectar uma relação entre a variável independente (o programa) e as variáveis dependentes (Judd & Kenny, 1982)? Uma amostragem aleatória de tamanho suficiente e a utilização judiciosa das diferentes técnicas de análise estatística permitem garantir, em parte, a validade da conclusão estatística de uma avaliação. Esse tipo de validade baseia-se igualmente na fiabilidade dos instrumentos de medida.

Desenho da Pesquisa: experimentação provocada

O Quadro 4 apresenta os principais desenhos utilizados em pesquisa experimental, bem como os vieses que podem afetar os resultados (tais como são descritos e numerados nas tabelas anteriores). Ele deveria auxiliar os avaliadores na escolha do desenho que poderá conferir a maior validade possível a sua pesquisa.

Quadro 4 - Desenho da pesquisa experimental: experimentação provocada

DESENHOS EXPERIMENTAIS	VISES QUE PODEM INFLUENCIAR NA VALIDADE INTERNA	VISES QUE PODEM INFLUENCIAR NA VALIDADE EXTERNA
PRÉ-TESTE, PÓS-TESTE COM GRUPO CONTROLE ALEATÓRIO		
	<p>4. Mortalidade experimental</p>	<p>9. Contágio 10. Reações compensatórias 11. Desejo de agradar ao avaliador 13. Intervenções compensatórias 14. Expectativa do experimentador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção 17. Interação entre as observações e a intervenção 18. Interação entre a seleção dos sujeitos e a intervenção</p>

DESENHOS EXPERIMENTAIS	VIÉS QUE PODEM INFLUENCIAR NA VALIDADE INTERNA	VIÉS QUE PODEM INFLUENCIAR NA VALIDADE EXTERNA
DESENHO DE SOLOMON COM QUATRO GRUPOS		
	4. Mortalidade experimental	9. Contágio 10. Reações compensatórias 11. Desejo de agradar ao avaliador 13. Intervenções compensatórias 14. Expectativa do experimentador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção 18. Interação entre a seleção dos sujeitos e a intervenção
PÓS-TESTE SOMENTE COM GRUPO CONTROLE ALEATÓRIO		
	4. Mortalidade experimental	9. Contágio 10. Reações compensatórias 11. Desejo de agradar ao avaliador 13. Intervenções compensatórias 14. Expectativa do experimentador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção 18. Interação entre a seleção dos sujeitos e a intervenção
ESTUDO ANTES - DEPOIS		
$O_1 \times O_2$	1. História 2. Maturação 3. Familiarização com o teste 7. Interações 8. Medida dos efeitos	12. Relação causal ambígua 14. Expectativa do experimentador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção 17. Interação entre as observações e a intervenção 18. Interação entre a seleção dos sujeitos e a intervenção
SÉRIE CRONOLÓGICA		
$O_1 \ O_2 \ O_1 \ X \ O_4 \ O_5 \ O_6$	1. História 8. Medida dos efeitos	11. Desejo de agradar ao avaliador 14. Expectativa do experimentador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção 17. Interação entre as observações e a intervenção 18. Interação entre a seleção dos sujeitos e a intervenção

DESENHOS EXPERIMENTAIS	VIÉS QUE PODEM INFLUENCIAR NA VALIDADE INTERNA	VIÉS QUE PODEM INFLUENCIAR NA VALIDADE EXTERNA
SÉRIE CRONOLÓGICA COM EXPERIMENTAÇÃO REPETIDA		
$X_a O_1 X_b O_2 X_c O_4 X_b O_5$	Nenhum	11. Desejo de agradar ao avaliador 14. Expectativa do experimentador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção 17. Interação entre as observações e a intervenção 18. Interação entre a seleção dos sujeitos e a intervenção
SÉRIE CRONOLÓGICA COM GRUPO CONTROLE NÃO EQUIVALENTE		
$\frac{O_1 O_2 O_3 X O_4 O_5 O_6}{O_1 O_2 O_3 O_4 O_5 O_6}$	Nenhum	9. Contágio 11. Desejo de agradar ao avaliador 14. Expectativa do experimentador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção 17. Interação entre as observações e a intervenção 18. Interação entre a seleção dos sujeitos e a intervenção
PRÉ-TESTE, PÓS-TESTE COM GRUPO CONTROLE NÃO EQUIVALENTE		
$\frac{O_1 X O_2}{O_1 O_2}$	5. Regressão para a média 7. Interações	9. Contágio 10. Reações compensatórias 11. Desejo de agradar ao avaliador 13. Intervenções compensatórias 14. Expectativa do experimentador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção 17. Interações entre as observações e a intervenção 18. Interação entre a seleção dos sujeitos e a intervenção
EXPERIMENTAÇÃO CONTRABALANCEADA OU QUADRADO LATINO		
Grupo1 $X_a O_1 X_b O_2 X_c O_3 X_d O_4$ Grupo2 $X_b O_1 X_d O_2 X_a O_3 X_c O_4$ Grupo3 $X_c O_1 X_a O_2 X O_3 X_b O_4$ Grupo4 $X_d O_1 X_c O_2 X_b O_3 X_a O_4$	7. Interação	9. Contágio 10. Reações compensatórias 11. Desejo de agradar ao avaliador 13. Intervenções compensatórias 14. Expectativa do experimentador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção 17. Interação entre as observações e a intervenção 18. Interação entre a seleção dos sujeitos e a intervenção

DESENHOS EXPERIMENTAIS	VIESSES QUE PODEM INFLUENCIAR NA VALIDADE INTERNA	VIESSES QUE PODEM INFLUENCIAR NA VALIDADE EXTERNA
CICLO INSTITUCIONAL COM ANÁLISE TRANSVERSAL - LONGITUDINAL		
Grupo 1 $X O_2 O_3 O_4$ Grupo 2 $O_2 X O_3 O_4$ Grupo 3 $O_3 O_4$	A validade do projeto para cada grupo é baixa; a validade global do projeto resulta da coerência dos resultados obtidos para os diferentes grupos	11. Desejo de agradar ao avaliador 14. Expectativa do experimentador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção 18. Interação entre a seleção dos sujeitos e a intervenção
REGRESSÃO DESCONTÍNUA		
Os sujeitos são divididos entre os dois grupos segundo uma regra conhecida mas não aleatória	3. Familiarização com o teste 4. Mortalidade experimental	15. Expectativa do experimentador 16. Interação entre os diferentes componentes de uma intervenção 18. Interação entre a seleção dos sujeitos e a intervenção

Leitura

R: Randomização dos sujeitos estudados entre grupos controle e experimentais.

X_2 : Intervenção, isto é, variável independente, (a) indica a modalidade da intervenção.

O_1 : Observação da ou das variáveis dependentes no período 1.

- : O grupo controle não é criado por randomização.

A comparação antes-depois

Quando se quer conhecer, por exemplo, os efeitos de um programa de redução do tabagismo, começa-se por medir o consumo de tabaco em um determinado grupo, implementa-se o programa e mede-se novamente o consumo de tabaco. Esse desenho é anotado como se segue:

$$O_1 X O_2$$

em que X representa a intervenção e O , as observações.

Os efeitos correspondem à diferença entre a observação realizada antes do início do programa e aquela que é realizada ao seu fim. Esse desenho não permite uma comparação perfeita, pois algumas hipóteses rivais podem ameaçar a validade dos resultados. Cinco vieses possíveis são enumerados.

O *viés de história*: um evento pode ocorrer entre as duas observações e levar a uma modificação (neste caso, do uso do tabaco) entre as pessoas do grupo observado. Quanto mais longo for o lapso de tempo entre as duas observações, maior será a possibilidade do surgimento de um viés de história. Inversamente, quanto mais curto for o hiato, menos provável será esse viés.

O *viés de maturação*: o tempo decorrido entre as duas observações pode provocar mudanças “naturais” entre os sujeitos, que envelhecem, ficam mais cansados ou menos atentos. Para tomar outro exemplo além daquele da redução do tabagismo, ao se medir o quociente intelectual dos alunos de uma turma antes e depois de uma aula, é possível que se obtenham escores menos elevados após a aula. Evidentemente, essa baixa não reflete uma diminuição da inteligência que seria atribuível à aula, mas ao cansaço.

O *viés de familiarização*: os sujeitos podem se acostumar a uma observação repetida. Por exemplo, responde-se geralmente melhor a uma pergunta quando esta já foi feita duas ou três vezes.

O *viés de medida dos efeitos*: os processos, os instrumentos de medida ou os observadores podem mudar.

O *viés de regressão para a média*: quando se escolhem sujeitos extremos (os melhores ou os menos bons), as observações subseqüentes se aproximam naturalmente da média.

A comparação entre dois grupos

No exemplo do programa de redução do tabagismo, é possível igualmente comparar dois grupos no mesmo momento: aquele que se beneficiou da intervenção e outro que não se beneficiou. Esse desenho é anotado como se segue:

$$\begin{array}{c} X O_1 \\ O_2 \end{array}$$

Os efeitos serão medidos pela diferença entre as duas observações (aqui, os dois grupos).

Utilizando-se ainda exemplo do programa de redução do tabagismo, podem-se selecionar dois grupos e oferecer o programa a apenas um dos dois. No mesmo momento, após a intervenção, mede-se a diferença de consumo de tabaco nos dois grupos. Esse desenho permite eliminar os vieses de história, de familiarização, de medida ou de regressão para a média. Ao contrário, caso se observe uma diferença no consumo de cigarros entre os dois grupos, não se deve necessariamente atribuí-la à intervenção. De fato, os dois grupos talvez tivessem níveis de consumo de tabaco diferentes antes da intervenção. Por isso, é possível que tenha ocorrido o que se chama de *viés de seleção*.

Além disso, mesmo que os grupos fossem comparáveis no início, uma *atrição seletiva* pode torná-los diferentes. Fala-se, então, em *viés de mortalidade*

experimental. É possível que os sujeitos que fazem parte do grupo experimental, os fumantes, por exemplo, se recusem a responder ao questionário por razões associadas à intervenção.

Finalmente, os dois grupos podem ter vivenciado acontecimentos diferentes. Fala-se, então, de viés de interação (seleção/história); ou eles evoluíram em ritmos diferentes (seleção/maturação).

A combinação dos desenhos

É possível limitar os vieses combinando-se os dois desenhos precedentes e medindo-se os efeitos antes e depois da intervenção, em dois grupos diferentes. Essa combinação de desenhos se lê da seguinte maneira:

$$\begin{array}{ccc} O_1 & X & O_2 \\ O_3 & & O_4 \end{array}$$

Os efeitos correspondem à diferença de mudança:

$$(O_2 - O_1) - (O_4 - O_3).$$

Esse desenho permite evitar os vieses de seleção, mesmo que os dois grupos sejam diferentes. Com efeito, como o que se mede é a diferença entre as mudanças operadas pelos dois grupos, o fato de os dois grupos serem diferentes no início não é importante. Por outro lado, a combinação dos desenhos permite evitar os vieses de história, de maturação, de medida dos efeitos ou de familiarização. Com efeito, os eventos se produzem nos dois grupos e são, portanto, controlados.

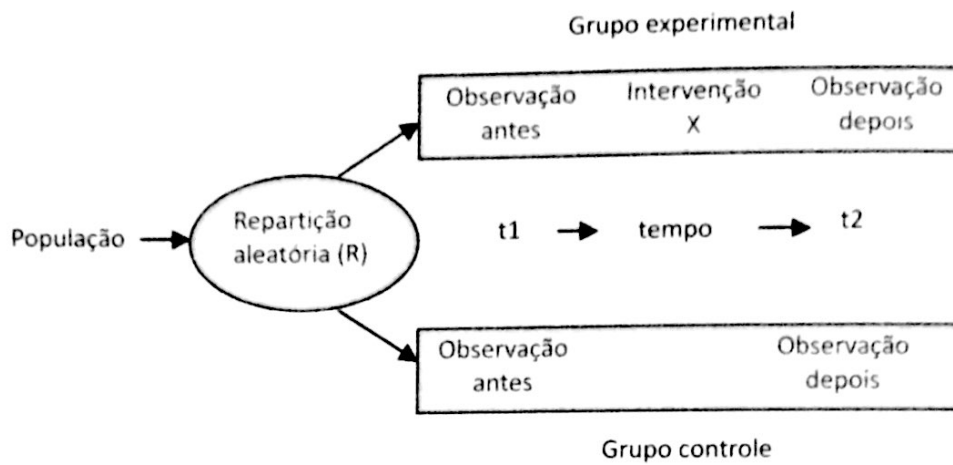
Em contraposição, a combinação dos desenhos encerra outros tipos de problemas que ameaçam a validade. Assim, pode ocorrer um viés de interação seleção/intervenção. Mais precisamente, é possível que algumas características positivas da população a quem o programa foi direcionado expliquem os seus efeitos. Nesse caso, é realmente a intervenção que causa uma diferença, mas a interação com as características favoráveis do grupo que se beneficia da intervenção amplia essa diferença. Esse desenho pode também comportar vieses de mortalidade experimental, de interação seleção/maturação ou de seleção/medida dos efeitos.

A repartição aleatória

A repartição aleatória da população estudada entre os grupos experimentais e os grupos controle é o único modo de se certificar de que os dois grupos

são exatamente comparáveis. A população estudada pode ser uma amostra, aleatória ou não, de uma população mais ampla.

Figura 3 - Modelo tradicional da pesquisa experimental



O fato de se escolher uma amostra ao acaso permite em seguida inferir resultados obtidos para o conjunto da população escolhida e melhorar a validade externa da avaliação. Porém, mesmo que os sujeitos-alvo da avaliação não constituam uma amostra aleatória de uma população, o fato de dividi-los ao acaso entre os dois grupos estudados permite assegurar uma efetiva validade interna à avaliação. O desenho de estudo ilustrado na Figura 3 é o modelo tradicional utilizado em pesquisa experimental. O único viés que talvez possa ameaçar a validade interna é a mortalidade experimental. Mas esse desenho comporta numerosos vieses (ver o Quadro 4) que podem ameaçar a validade externa (Contandriopoulos *et al.*, 1990).

Diversos desenhos de pesquisa são possíveis em uma situação de experimentação provocada: o presente capítulo apresentou os mais comuns. Como vimos, existem vários vieses que podem afetar a validade interna ou externa de uma pesquisa. Os três vieses a que um avaliador deveria ficar particularmente atento são os vieses de história, de maturação e de seleção. Os vieses de história e de seleção comportam um risco fundamental, pois impedem que se saiba se a intervenção avaliada se encontra realmente na origem dos efeitos observados.

Desenho de Pesquisa: experimentação invocada

Em muitos casos, não é possível utilizar desenhos de experimentação provocada, e o pesquisador deve se valer das variações naturais para avaliar os efeitos de uma intervenção (experimentação invocada). Ele procura com-

parações possíveis sem manipular os grupos existentes. Mais uma vez, a ideia consiste em comparar grupos entre si ou em um mesmo grupo em momentos diferentes. Os desenhos são os mesmos da experimentação provocada, sem randomização, evidentemente. É plenamente possível ter em mente alguns “antes-depois” com base nos dados sobre a população antes que ela seja exposta à intervenção. Pode-se igualmente comparar grupos entre si escolhendo-se um grupo que foi exposto e um grupo controle que não foi exposto à intervenção.

O Quadro 5 apresenta alguns desenhos de experimentação invocada e os principais vieses que lhes são associados.

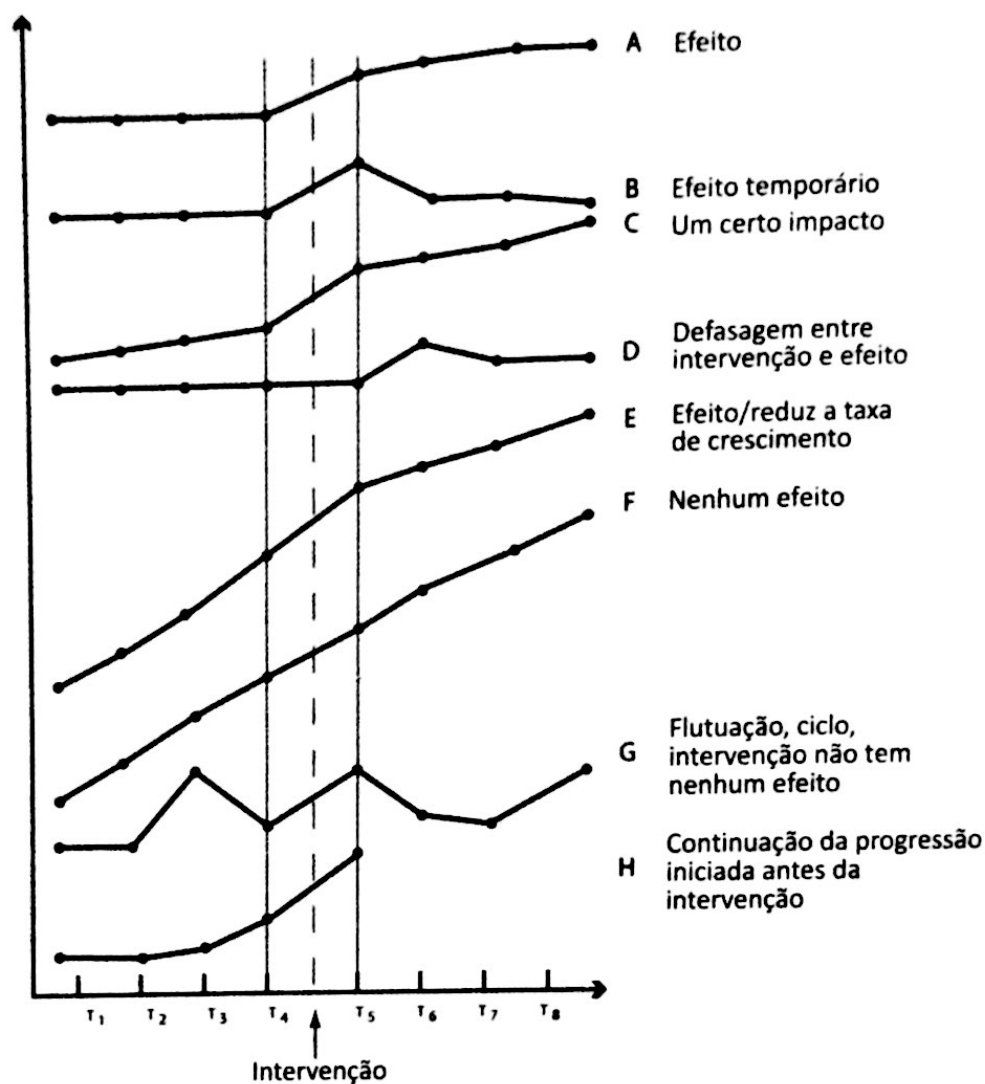
O que, frequentemente, é muito difícil quando das experimentações invocadas é encontrar dados disponíveis. Quando esse acesso é facultado, é possível sem dificuldade fazer múltiplas observações no decorrer do tempo. Pode-se, então, escolher um desenho em que se acompanha uma população ou um fenômeno ao longo do tempo. Esses desenhos chamados séries cronológicas ou séries temporais múltiplas são muito interessantes, pois têm uma validade muito boa. Um exemplo permite compreender melhor este ponto. Consideremos a hipótese de que um governo adotou, por exemplo, em 1º de julho de 2006, uma nova política em matéria de acesso à condução de automóveis: antes de obter sua carteira de habilitação, um novo motorista deve ter sido detentor de uma carteira de aprendiz durante um ano. Para saber se essa medida é eficaz, é preciso começar por construir o modelo lógico teórico da intervenção. É provável, então, que seja consensual que essa medida tem por objetivo reduzir a mortalidade e a morbidade associadas aos acidentes de trânsito. A medida dos efeitos poderia ser, simplesmente, a taxa de acidentes, visto que a hipótese causal que postula uma relação entre acidentes e mortalidade é robusta. É possível também escolher a taxa de acidentes mortais, se os dados estiverem disponíveis. Ora, na maioria das jurisdições há registros semanais e mensais sobre os acidentes. Por isso, é possível conhecer o número de acidentes fatais em uma base mensal. Para avaliar o efeito da nova política, basta analisar a tendência temporal dos acidentes mortais por mês desde, por exemplo, 2002, e verificar se essa tendência se modifica a partir de julho de 2007 (para considerar o período de latência ligado à implantação do programa). Esse tipo de desenho, que utiliza dados em série, é muito eficaz, pois enquanto o “antes-depois” mede apenas uma variação estabelecendo a diferença entre os resultados em dois determinados momentos, os estudos seriais permitem analisar tendências. Eles permitem, desse modo, evitar diversos vieses.

Quadro 5 – Desenho da pesquisa experimental: experimentação invocada

ESTUDOS COMPARATIVOS EX POST	VIESES QUE PODEM INFLUENCIAR NA VALIDADE INTERNA	VIESES QUE PODEM INFLUENCIAR NA VALIDADE EXTERNA
PÓS-TESTE SOMENTE		
$X O_2$	1. História 2. Maturação 3 Familiarização com o teste 4. Mortalidade experimental 6. Seleção	12. Relação causal ambígua 14. Expectativa do experimentador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção 18. Interação entre a seleção dos sujeitos e a intervenção
PÓS-TESTE SOMENTE COM GRUPO CONTROLE EQUIVALENTE		
$X O_2$ — O_2	2. Maturação 4. Mortalidade experimental 6. Seleção 7. Interações	12. Relação causal ambígua 14. Expectativa do experimentador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção 18. Interação entre a seleção dos sujeitos e a intervenção
ESTUDO ANTES - DEPOIS		
$O_1 X O_2$	1. História 2. Maturação 3 Familiarização com o teste 7. Interações 8. Medida dos efeitos	12. Relação causal ambígua 14. Expectativa do experimentador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção 18. Interação entre a seleção dos sujeitos e a intervenção
PRÉ-TESTE, PÓS-TESTE COM GRUPO CONTROLE NÃO EQUIVALENTE		
$O_1 X O_2$ — $O_1 O_2$	4. Mortalidade experimental 5. Regressão para a média 7. Interações	11. Desejo de agradar ao avaliador 14. Expectativa do experimentador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção 18. Interação entre a seleção dos sujeitos e a intervenção
ESTUDOS SERIAIS	VIESES QUE PODEM INFLUENCIAR NA VALIDADE INTERNA	VIESES QUE PODEM INFLUENCIAR NA VALIDADE EXTERNA
Séries cronológicas $O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8$	1. História 8. Medida dos efeitos	11. Desejo de agradar ao avaliador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção

ESTUDOS SERIAIS	VIESES QUE PODEM INFLUENCIAR NA VALIDADE INTERNA	VIESES QUE PODEM INFLUENCIAR NA VALIDADE EXTERNA
Séries cronológicas $O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8$	1. História 8. Medida dos efeitos	11. Desejo de agradar ao avaliador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção
Séries cronológicas múltiplas	Validade interna muito grande	11. Desejo de agradar ao avaliador 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção 18. Interação entre a seleção dos sujeitos e a intervenção
Estudos casos controle	Validade dependente do processo de seleção dos casos e dos controles	12. Relação causal ambígua 16. Interação entre os diferentes componentes de uma intervenção
ESTUDOS CORRELACIONAIS	VIESES QUE PODEM INFLUENCIAR NA VALIDADE INTERNA	VIESES QUE PODEM INFLUENCIAR NA VALIDADE EXTERNA
Investigação estatística das relações entre uma variável dependente e diversas variáveis independentes de modo a isolar a influência daquela que é objeto da hipótese de pesquisa	Validade assegurada pela conformidade entre o modelo teórico e as abordagens estatísticas escolhidas	12. Relação causal ambígua 15. Interação entre a intervenção e a situação experimental 16. Interação entre os diferentes componentes de uma intervenção

Figura 4 - Diferentes trajetórias de variação dos efeitos



Fonte: adaptado de Campbell & Stanley (1963), fig. 5, p. 35.

A Figura 4 ilustra várias situações observáveis. Na situação A, uma análise antes-depois da implantação política (diferença entre os tempos T₄ e T₅) teria mostrado que a intervenção gerou um efeito. Os dados cronológicos mostram isso também. Na situação B, o efeito é apenas temporário, o que um desenho antes-depois não teria permitido observar. A situação C permite constatar que a intervenção acelerou a tendência, que já era de alta, mas que o ritmo de crescimento foi retomado quase imediatamente após a intervenção. Na situação D, o estudo cronológico mostra que há uma defasagem entre o momento em que a intervenção é oferecida e o momento em que os efeitos são obtidos. Uma análise antes-depois teria permitido concluir que o efeito da intervenção é nulo. O exemplo F ilustra a situação inversa. Um estudo antes-depois teria indicado

um efeito, quando na verdade o estudo serial mostra que a intervenção não teve efeito, já que a tendência temporal não é afetada pela intervenção. Na situação G, observa-se uma flutuação do indicador de efeitos. A intervenção não parece ter tido impacto. Um estudo antes-depois, no entanto, teria indicado um efeito positivo. Finalmente, a situação H mostra que a intervenção acelerou uma tendência já presente. O efeito da intervenção não é, portanto, tão grande quanto aquele que um estudo antes-depois teria revelado.

O único viés que poderia reduzir a validade interna das séries cronológicas é o viés de história. Portanto, é preciso certificar-se de que acontecimentos que poderiam alterar os efeitos não aconteçam ao mesmo tempo que a intervenção. Para retomar o exemplo de uma política que teria por objetivo impor aos jovens motoristas ter acesso apenas a uma carteira de aprendiz durante um ano, seria necessário certificar-se, para que se possa atribuir a redução dos acidentes de trânsito à política, de que no mesmo momento o governo não tenha reforçado os controles de velocidade. Para eliminar esse viés de história, sempre é possível verificar se algo aconteceu simultaneamente à intervenção. Outra possibilidade seria encontrar uma população de comparação para fazer um estudo cronológico comparativo.

Exemplo

Análise dos efeitos do programa Nascer Iguais – Crescer com Saúde

BRODEUR, J.-M. *et al.* Le programme québécois 'Naître égaux - Croître en santé': études des effets sur la santé des mères et nouveaux-nés. *Santé, Société et Solidarité*, 3(1): 119-127, 2004.

O programa Nascer Iguais – Crescer com Saúde (já tratado no capítulo 5) se dirige às mulheres com pouca escolaridade e renda familiar baixa. Tem por objetivo reduzir a taxa de nascimento de bebês cujo peso é inferior a 2.500 g, reduzir a proporção de crianças que apresentam atraso de crescimento intrauterino, reduzir a taxa de nascimentos prematuros, a incidência dos abusos e da negligência em relação às crianças. A partir da vigésima semana de gravidez, as participantes do programa recebiam suplementos alimentares e um acompanhamento personalizado.

O objetivo do projeto de avaliação era determinar se a fase pré-natal do programa tinha influência significativa sobre a saúde das mães e dos recém-nascidos desfavorecidos.

As participantes do projeto (n=1.340) foram divididas aleatoriamente em dois grupos: um grupo que se beneficiava do programa (grupo experimental) e um grupo a quem se forneciam apenas suplementos alimentares (grupo controle). A hipótese principal era de que haveria menos mulheres tendo bebês com peso inferior a 2.500 g entre as mulheres que se beneficiam do programa do que entre aquelas do grupo controle.

O recrutamento das participantes ocorria em dez territórios de Centros Locais de Serviços Comunitários (CLSC) por meio de correspondências, cartazes e fôlderes, e da promoção do programa entre os agentes da intervenção. Diferentes critérios de inclusão foram elaborados. Para ser admissível no programa, era preciso estar entre 16 e 22 semanas de gestação, ter completado menos de 11 anos de escolaridade, viver em um lar de baixa renda, compreender e se expressar em francês ou em inglês, morar no território de um CLSC participante. As mulheres que tinham algum problema psiquiátrico agudo diagnosticado, alguma deficiência intelectual diagnosticada ou que esperavam gêmeos estavam excluídas do estudo.

Os dados foram coletados mediante a administração de questionários, a realização de entrevistas no pré e no pós-natal e consulta aos prontuários de acompanhamento e de hospitalização. Análises estatísticas foram realizadas.

Os resultados mostraram que as mulheres que se beneficiaram do programa adquiriram claramente maior apoio social. O programa teve um efeito protetor contra os sintomas depressivos pós-parto. As participantes dos dois grupos apresentavam indicadores de saúde similares para o período pré-natal, para o parto ou para o período pós-parto imediato. As participantes do programa tinham muitas vezes menos anemia no pós-natal do que as mulheres do grupo controle. O programa teve como efeito a melhora dos hábitos alimentares das mulheres. As mulheres nascidas no Canadá e que participavam do programa amamentaram de bom grado seu bebê. Finalmente, as mulheres que não amamentavam ofereciam a seu bebê mais preparações lácteas do que leite de vaca.

O programa, todavia, não apresentou efeito sobre a insuficiência de peso no nascimento, nem sobre o atraso do crescimento intrauterino e a prematuridade.

Quem emite um juízo avalia; por isso, a avaliação é universal. Mas essa atividade humana pode ser mais ou menos formalizada, e estratégias podem ser adotadas para aumentar a sua validade. Na prática, o rigor metodológico tanto quanto a resposta às necessidades reais dos organismos solicitantes e a preocupação constante com o uso da informação produzida são as melhores garantias de uma avaliação bem-sucedida.

Exaustiva e pragmática, esta obra apresenta o modelo de avaliação que há vinte anos tem se elaborado em torno do curso "Métodos de avaliação" do setor da saúde pública da Universidade de Montreal. A maioria dos exemplos utilizados provém do campo da saúde, mas o modelo proposto se aplica ao estudo das intervenções de todos os setores, como a educação, a administração ou as ciências políticas.

Este livro permitirá aos estudantes, aos pesquisadores e aos profissionais que recorrem à avaliação definir melhor as intervenções e escolher as abordagens e as perguntas mais bem adaptadas a cada contexto.

Os organizadores ensinam avaliação de programa e são, todos, originalmente membros do Grupo de Pesquisa Interdisciplinar em Saúde (Gris) da Universidade de Montreal. Médicos, profissionais da saúde, pesquisadores e docentes, os presentes autores estão entre os melhores especialistas da avaliação.