

# Um modelo alternativo ao estudo experimental para inferir causalidade em avaliações do impacto de projetos sociais

Thomaz Chianca<sup>1</sup>

1 Possui Doutorado em Avaliação Interdisciplinar pela Universidade de Western Michigan nos EUA (2008), Mestrado em Saúde Pública pela Universidade da Carolina do Norte em Chapel Hill nos EUA (1996), Graduação em Odontologia pela Universidade Federal do Rio de Janeiro (1985). Trabalha como odontólogo na Faculdade de Odontologia da Universidade Federal do Rio de Janeiro (UFRJ) desde 1990 onde realiza pesquisas na área de saúde pública, ensina metodologia científica para alunos de mestrado, e supervisiona projetos de extensão com alunos de graduação. Atua também como avaliador de programas, projetos e políticas sociais no Brasil e em diversos países da América Latina de do Norte, Sudeste Asiático, África, e Leste Europeu.

## Resumo

---

ALGUNS ESTUDIOSOS DEFENDEM QUE APENAS MÉTODOS EXPERIMENTAIS OU QUASE EXPERIMENTAIS PODEM DETERMINAR O IMPACTO DAS MUDANÇAS PROVOCADAS POR UM PROJETO. ESSA PREMISSA É QUESTIONADA NESTE ARTIGO, TANTO EM RELAÇÃO À SUA VALIDADE, QUANTO NO QUE TANGE À SUA VIABILIDADE E À SUA UTILIDADE DE APLICAÇÃO NA PRÁTICA COTIDIANA DOS AVALIADORES. O PRESENTE TRABALHO DISCUTE OS FUNDAMENTOS LÓGICOS E AS ESTRATÉGIAS DE UMA METODOLOGIA NÃO EXPERIMENTAL DESENVOLVIDA PARA AVALIAR MAIS DE 200 PROJETOS FOCADOS NA REDUÇÃO DA POBREZA RURAL E NA PROTEÇÃO DO MEIO AMBIENTE, EM 23 PAÍSES. ALÉM DOS DETALHES SOBRE O MODELO AVALIATIVO DESENVOLVIDO, SERÃO APRESENTADAS AS LIÇÕES APRENDIDAS A PARTIR DA APLICAÇÃO DO MODELO, BEM COMO DEBATIDA SUA RELEVÂNCIA PARA O CAMPO DA AVALIAÇÃO, COMO ALTERNATIVA PLAUSÍVEL AO MODELO EXPERIMENTAL. TAMBÉM SERÃO MOSTRADAS AS ESTRATÉGIAS CRIATIVAS ADOTADAS PARA AVALIAR O IMPACTO DOS PROJETOS, EM UM CONTEXTO NO QUAL HAVIA LIMITAÇÕES SIGNIFICATIVAS DE TEMPO, DE RECURSOS E DE ACESSO A INFORMAÇÕES.

## Abstract

---

*SOME ACADEMICS SUPPORT THE IDEA THAT ONLY EXPERIMENTAL AND SOME QUASI-EXPERIMENTAL DESIGNS CAN DETERMINE THE EXTENT TO WHICH A PROJECT CAUSES CHANGES. THIS ASSUMPTION IS QUESTIONED IN THIS PAPER ON THE BASIS OF ITS VALIDITY AND ALSO IN RELATION TO ITS FEASIBILITY AND UTILITY IN REAL-LIFE SETTINGS. THIS PAPER DISCUSSES THE LOGICAL FOUNDATIONS AND STRATEGIES OF A NON-EXPERIMENTAL METHODOLOGY DEVELOPED TO EVALUATE MORE THAN 200 PROJECTS FOCUSED ON RURAL POVERTY REDUCTION AND ENVIRONMENTAL PROTECTION IN 23 COUNTRIES. BESIDES DESCRIBING THE EVALUATION MODEL, LESSONS LEARNED FROM THE APPLICATION OF THE MODEL WILL BE PRESENTED, AS WELL AS ITS RELEVANCE TO THE EVALUATION FIELD WILL BE ASSESSED AS A PLAUSIBLE ALTERNATIVE TO THE EXPERIMENTAL MODEL. IT WILL ALSO COVER THE CREATIVE STRATEGIES DEVELOPED TO EVALUATE IMPACT WITHIN A CONTEXT OF SIGNIFICANT LIMITATIONS OF TIME, RESOURCES AND ACCESS TO INFORMATION.*

### PALAVRAS-CHAVE:

Avaliação de programas; Avaliação de impacto; Causalidade.

## Introdução

---

Quase todas as avaliações de programas sociais precisam responder, entre outras perguntas, sobre o quanto esses programas contribuíram para que a vida de indivíduos, famílias, grupos ou o contexto no qual as pessoas vivem fosse mudado. Portanto, o conceito de causalidade é um aspecto essencial para tais processos avaliativos. Uma das questões centrais que o avaliador precisa considerar é a extensão da responsabilidade dos programas em relação às mudanças observadas. Para poder responder apropriadamente a esta questão é necessário não só encontrar evidências que conectem o programa com essas transformações, como também excluir de maneira sistemática as principais alternativas plausíveis que também possam ter causado as alterações. Por exemplo, o aumento da renda familiar observado foi devido à capacitação oferecida pelo projeto? As mudanças introduzidas no processo produtivo resultaram dessa capacitação ou do fato de que membros da família conseguiram empregos com melhor remuneração ou ainda devido à introdução de programas nacionais ou regionais para complementação de renda de famílias mais pobres?

Este trabalho faz uma análise crítica da ideia de que os estudos randomizados controlados são o “padrão ouro” para determinar o impacto causado por uma intervenção. Discute-se também uma alternativa para produzir evidências robustas de que as ações realizadas por um programa social causam determinados resultados na sua população-alvo. Essa alternativa está baseada em metodologia não experimental desenvolvida para avaliar mais

de 200 projetos focados na redução da pobreza rural e na proteção do meio ambiente, apoiados por uma organização não governamental internacional, em 23 países em desenvolvimento, na América Latina, Ásia, África e leste europeu.

## Desconstruindo o paradigma

---

Quando se estuda o assunto um pouco mais a fundo, chama a atenção a diversidade das definições do que é considerado como “avaliação do impacto de programas” a partir de diferentes referenciais. No senso comum do campo social, o impacto de um programa é visto como “quaisquer mudanças produzidas por aquela intervenção”. Levando-se em conta definições mais amplas, como a proposta pelo Comitê de Assistência ao Desenvolvimento da Organização Europeia de Cooperação para o Desenvolvimento (UNEG, 2013), seria possível complementar essa definição afirmando que as mudanças podem ser tanto positivas quanto negativas, esperadas ou não e direta ou indiretamente influenciadas pelo programa.

Entretanto, no campo econômico e em outras áreas afins, a definição se dá a partir de uma fórmula matemática:  $Y1 - Y0 = \text{Impacto}$ . Traduzindo: é a diferença no indicador de interesse (Y), com a intervenção (Y1) e sem a intervenção (Y0), conhecido também como contrafactual. Esse método é defendido por boa parte dos integrantes do Banco Mundial e por organizações de cooperação bilateral como a USAID. Por meio dele, assume-se

que só é possível fazer qualquer atribuição de causa, de forma rigorosa, se houver uma comparação com o contrafactual. Este, quase sempre, inclui a criação de um grupo controle, composto por membros que não foram expostos ao programa ou que foram submetidos a outro tipo de intervenção, mas que possuem características similares às dos grupos que participaram do programa.

Thomas Cook e Michael Scriven (Cook et al. 2010) discutiram em profundidade esse tema em um seminário realizado na Universidade de Western Michigan (WMU) em 2008. Ambos são referência internacional no campo da avaliação e possuem visões distintas em relação à determinação da causalidade. Cook é cientista social e referência acadêmica no uso de modelos experimentais e quase experimentais para inferência causal generalizada (SHADISH, COOK & CAMPBELL, 2002). Scriven é matemático e filósofo, um dos primeiros a criar bases fundamentais e a publicar trabalhos sobre a lógica e a metodologia de avaliação (SCRIVEN, 1967). Ele tem publicado a respeito da determinação da causalidade desde os anos 1960 (SCRIVEN, 1966).

A seguir, apresenta-se uma interpretação dos principais argumentos que surgiram durante o seminário da WMU, questionando alguns dos pilares que sustentam a ideia de que os estudos randomizados controlados devem

ser o método de escolha para a avaliação de impacto.

## **CONTRAFACTUAL NÃO É FATOR ESSENCIAL E NECESSÁRIO PARA SE INFERIR CAUSALIDADE**

Impacto pode ser observado, não precisa ser testado sempre. Quem pode negar que o barulho que resulta de um bater de palmas não foi causado pela ação de duas mãos se chocando uma com a outra? Em uma comunidade sem acesso a água encanada, as famílias (principalmente mulheres e crianças) gastavam, em média, três horas por dia para trazer água de um poço para casa. Depois que uma organização instalou um sistema de água e saneamento na comunidade, essas famílias não tiveram mais que cumprir essa tarefa. Quem pode negar que as três horas ganhas por não precisarem mais buscar água foram resultado do projeto realizado? E a provável diminuição de diarreia crônica em longo prazo? Fica óbvio, nesses casos, que não é necessário um grupo controle para fazer esse tipo de inferência causal.

Há um outro argumento, mais filosófico, que refuta a necessidade do contrafactual para determinar causa: é a questão da sobre-determinação (SCRIVEN 2008). São os casos em que um número de fatores está presente no contexto, e qualquer um deles poderia produ-

zir o mesmo resultado, caso isso não ocorresse devido à intervenção.

## **RANDOMIZAÇÃO NÃO ELIMINA TODAS AS CAUSAS ALTERNATIVAS**

O primeiro problema em estudos experimentais envolvendo pessoas é que, por razões éticas, é necessário informar que essas pessoas farão parte de um experimento. A não ser em casos mais raros, como em estudos farmacológicos, nos quais é relativamente fácil dissimular a droga ativa e o placebo, as pessoas rapidamente identificam quem está recebendo o potencial benefício do programa e quem não está. Pelo menos três problemas aparecem nessa situação que fazem com que a intervenção não seja a única diferença entre os grupos:

- I. **atrição diferencial:** quando um alto número de pessoas do grupo controle resolve deixar o estudo, chegando a um nível crítico que interfere na qualidade das comparações estatísticas;
- II. **contaminação cruzada:** quando os participantes do grupo experimental e do controle trocam informação, e o segundo começa a fazer, em todo ou em parte, o que o experimental está fazendo;
- III. **efeito Hawthorne:** quando as pessoas mudam seu comportamento positivamente por saberem que estão recebendo a intervenção, ou negativamente, quando não estão recebendo a intervenção.

É necessário ter um grupo de pesquisadores altamente competente para conseguir con-

trolar ou minimizar esses problemas. Uma observação interessante, neste caso, é o fato de que, para identificar e tentar solucionar esses problemas, é preciso ter habilidades de investigação interativa e em profundidade com as pessoas envolvidas, sendo que essas habilidades vêm do campo qualitativo. Observa-se assim que, no coração do método quantitativo, é fundamental o uso de métodos qualitativos.

## **ESTUDOS RANDOMIZADOS SÃO RESTRITOS**

Outro argumento, apresentado por Cook (2010), um dos pais dos estudos quantitativos, é o fato de que os estudos randomizados controlados só conseguem responder a um número restrito de perguntas fora do campo causal. E as perguntas causais estão sempre relacionadas a variáveis passíveis de manipulação e quase nunca conseguem explicar por que os resultados ocorreram.

A questão fundamental aqui é que não se pode deixar que um método de pesquisa determine as perguntas avaliativas a serem feitas. Primeiro são encaminhadas as perguntas, e depois são determinados os métodos que melhor servirão para respondê-las de forma consistente e com a precisão requerida pelo contexto. Se o método de estudo randomizado controlado é colocado como padrão ouro para os estudos de impacto, então as perguntas avaliativas estarão limitadas a poucas variáveis passíveis de mensuração.

## **ESTUDOS EXPERIMENTAIS NÃO CONSEGUEM DAR CONTA DOS IMPACTOS NÃO ESPERADOS, PRINCIPALMENTE OS NEGATIVOS**

Complementarmente ao ponto anterior, uma vez que os estudos experimentais só olham para variáveis predeterminadas e manipuláveis, qualquer coisa que ocorra como resultado da intervenção e que não esteja na lista prefixada de variáveis ficará fora do radar. É sempre necessária a inclusão de métodos qualitativos para explorar os vários efeitos colaterais de intervenções sociais. Esses efeitos, algumas vezes, podem ser sérios o suficiente para anular a importância ou o valor de qualquer impacto positivo que venha a ser observado nas variáveis de interesse.

## **ESTUDOS RANDOMIZADOS CONTROLADOS NÃO SÃO OS ÚNICOS QUE PODEM PRODUZIR CONCLUSÕES CONVINCENTES SOBRE CAUSA E EFEITO**

Há vários modelos quantitativos que receberam uma denominação de aparente inferioridade – estudos quase experimentais –, como

a série temporal interrompida<sup>2</sup> (ST. CLAIR, COOK e HALLBERG, 2014) e a regressão descontínua<sup>3</sup> (IMBENS e LEMIEUX, 2008). Esses estudos, quando bem realizados, apresentam conclusões sobre relações causais tão bem fundamentadas quanto um bom experimento. Várias outras abordagens baseadas em modelos teóricos e qualitativos podem chegar a conclusões de causalidade de forma convincente, dependendo do nível de precisão que se deseja. Jane Davidson (2005) identificou algumas dessas abordagens e recomendou a utilização combinada delas para aumentar o nível de confiabilidade/convencimento das conclusões (detalhes sobre essas abordagens podem ser encontrados em artigo sobre causalidade em avaliação (CHIANCA, 2012)):

- Identificar os mecanismos causais subjacentes através do modelo lógico.
- Checar se o momento em que os resultados ocorreram está dentro de um tempo factível.
- Checar se a “dose” da intervenção justifica o “efeito” observado.
- Perguntar às pessoas que participaram do programa.

2 Quando a intervenção é realizada e interrompida por períodos distintos e há medição das variáveis de interesse em vários momentos para determinar se há algum efeito notável.

3 Quando, em estudos com número elevado de participantes, comparam-se sujeitos afetados diretamente pela intervenção de interesse com outros próximos a eles que não foram afetados diretamente pela intervenção (por exemplo, alunos premiados).

Um dos métodos usados com frequência é o modelo geral de eliminação (general elimination model (GEM)), desenvolvido por Michael Scriven (1991). Tal método consiste na criação de um “livro de causas prováveis” – além da intervenção em avaliação – para os efeitos que se pretende produzir e na realização de um trabalho minucioso para eliminar (ou não) as causas alternativas. O livro de causas é constituído principalmente a partir da revisão cuidadosa de teorias publicadas sobre o assunto, de consultas a especialistas na área e de entrevistas em profundidade com informantes-chave que conhecem de perto o contexto e a área de conhecimento nos quais se insere a intervenção. Para eliminar as possíveis causas alternativas usa-se o método – também desenvolvido por Scriven (1991) – conhecido como *modus operandi*. Trata-se do trabalho do investigador metucioso que, a partir do conhecimento sobre a forma como cada causa alternativa atua no detalhe, busca evidências específicas para saber se o modo de operar esteve presente nas mudanças observadas.

## Aplicação do modelo geral de eliminação para inferir causalidade

---

O exemplo usado aqui vem de uma avaliação de impacto de projetos apoiados por uma ONG internacional que atua em vários países doando animais e realizando capacitações para ajudar famílias pobres que vivem em zonas rurais. O foco dos projetos é a melhoria das condições de vida das famílias e a pro-

teção do meio ambiente. Ao longo de sete anos foram realizadas avaliações externas do trabalho desta organização em 23 países, envolvendo mais de 200 projetos. A coordenação desse trabalho esteve a cargo de Michael Scriven, e o autor deste artigo atuou como gerente dessas avaliações externas.

Em um estudo de impacto dessa natureza existem pelo menos dois grandes problemas de desenho que precisavam ser tratados. O primeiro diz respeito à atribuição de impacto, e o segundo, a como determinar os impactos a serem estudados e as evidências que precisavam ser consideradas.

### ATRIBUINDO IMPACTO

Se fosse utilizado o “padrão ouro” defendido por alguns estudiosos – estudo experimental com alocação aleatória de sujeitos –, haveria a necessidade de sortear comunidades ou grupos de pequenos agricultores para receberem (grupo experimental) ou não (grupo controle) as ações do projeto. Entretanto, neste caso específico, como em muitos que ocorrem nas práticas dos avaliadores, foi demandada uma avaliação de projetos que já estavam em curso por alguns anos.

Este fato impossibilitava uma seleção aleatória dos projetos que seriam alocados para o grupo que não receberia o tratamento, mesmo que não houvesse razões éticas e econômicas que tornassem essa abordagem inviável. Isso quer dizer que o método considerado mais eficaz para determinar causalidade, ou seja,

os efeitos provocados pelas intervenções do projeto, estava excluído. Entretanto, como já mencionado, essa não é a única metodologia que consegue fazer inferências causais. Provou-se ser possível, tomando-se o devido cuidado, utilizar métodos alternativos para tal fim.

Os avaliadores buscaram desenvolver uma metodologia para dar conta do problema de inferência causal em projetos sociais que pudesse ser comparável, em termos de qualidade e rigor, aos melhores exemplos de estudos experimentais. Para tanto, partiu-se do princípio de que é possível inferir causalidade através do controle rigoroso da temporalidade (momento da intervenção e da observação de resultados) e de variáveis de confundimento (fatores que também poderiam explicar as mudanças observadas independentemente das ações implementadas pelo programa). Essa abordagem foi baseada no GEM – modelo proposto por Scriven.

Na adaptação feita, os avaliadores reconstruíram a linha de base dos projetos para chegar ao contrafactual a partir de informações secundárias sobre os projetos e de entrevistas com diversos informantes-chave, incluindo anciãos, participantes dos projetos e líderes locais e regionais. Procurou-se fazer essa reconstrução de forma ampla, incluindo diferentes atores, acessando toda a documentação disponível e buscando a triangulação de

informações para assegurar a maior confiabilidade possível nas estimativas realizadas. Porém, como em qualquer outro método de pesquisa, este também estava sujeito a vieses, principalmente relacionados à memória de alguns eventos passados.

A situação atual de cada local em relação aos diferentes critérios avaliativos – definidos pelos avaliadores e membros da organização contratante em consulta com representantes dos diferentes países – foi determinada a partir de observações da realidade durante as visitas aos projetos, de entrevistas com informantes-chave e membros das comunidades, e da revisão de documentação pertinente.

O segundo passo incluiu a identificação (criação do “livro das causas”) e eliminação sistemática (onde existisse evidência suficiente para isso) de todos os fatores que poderiam, de forma plausível, provocar os efeitos que foram observados ou inferidos.

Neste caso específico, levantou-se a presença ou não de vários fatores que poderiam ter causado ou contribuído de forma significativa para as mudanças observadas no contexto dos projetos em zonas rurais pobres, entre eles:

- Programas realizados pelos governos nacionais ou locais que tenham chegado às famílias participantes das ações do projeto.

- Apoio de outras organizações filantrópicas.
- Remessa de dinheiro por parte de membros dessas famílias vivendo em outras regiões ou países.
- Ganhos monetários conseguidos a partir de projetos comunitários de autoajuda ou de financiamentos coletivos não vinculados ao projeto.
- Aumento do preço do leite ou de outros subprodutos animais no mercado local.
- Benefícios cumulativos proporcionados por alguns anos de clima favorável para a lavoura ou por melhorias no suprimento de água para irrigação.
- Novas tecnologias que disponibilizaram no mercado alimentos ou remédios para animais ou serviços veterinários de melhor qualidade e a preço acessível.
- Sementes modificadas, disponíveis no mercado a preço acessível, que produziram melhor pasto para os animais.

Chegar a esta lista no contexto e na cultura dos países e comunidades incluídas neste estudo requereu um esforço considerável e sistemático por parte da equipe de avaliação. Porém, não exigiu nenhum método sofisticado de estudo experimental ou modelos estatísticos complicados.

Esta análise altamente contextual só foi possível devido à possibilidade de acesso ao conhecimento coletivo dos anciãos das comunidades e de especialistas de instituições

de pesquisa em termos de conteúdo e localização geográfica dos projetos. Observações diretas, entrevistas em profundidade, além da revisão de documentos estratégicos sobre as comunidades e projetos, foram as principais fontes de evidência sobre o sucesso ou não das intervenções realizadas.

Os projetos sociais analisados eram complexos, e a cadeia de inferências que levou a equips de avaliação a chegar a conclusões sobre os impactos variou consideravelmente entre os países, entre os projetos dentro de um mesmo país e, algumas vezes, até entre diferentes grupos ou regiões de um mesmo projeto. Tal cadeia pode ir desde observações diretas simples até redes complexas de inferências conectadas às variáveis dependentes mais intrincadas, como as ligadas às mudanças nos valores e nas atitudes que dependem dos padrões éticos de cada cultura. Os avaliadores externos que realizaram as visitas de campo participaram de formações específicas de capacitação para realizar as inferências causais mais longas e complexas, a fim de se chegar a conclusões avaliativas válidas e relevantes.

## **QUE IMPACTOS CONSIDERAR E QUE EVIDÊNCIAS BUSCAR**

O segundo problema enfrentado nesse processo avaliativo envolveu a gama de impactos que deveria ser considerada num estudo como esse, que evidências precisavam ser coletadas e quais as estratégias mais adequadas para coletá-las.

A definição convencional de impacto está focada principalmente em efeitos positivos em termos econômicos, de saúde e/ou de educação. São esses, sem dúvida, os efeitos mais esperados e que podem ser mais facilmente observados. Entretanto, eles não são os únicos (ou efeitos potenciais) importantes que um avaliador deveria se preocupar em buscar.

Para facilitar o entendimento da abordagem, utilizou-se uma metáfora que envolve a criação de pegadas e outras mudanças no ambiente causadas por um animal enquanto caminha. Os impactos mais visíveis são aqueles deixados pelo animal diretamente sobre a superfície do solo. Estes representariam ganhos materiais mais aparentes, como o aumento na renda, no número de animais possuídos, no tempo que os filhos permanecem na escola, ou melhorias na saúde dos membros das famílias. Algumas vezes tais mudanças não são tão aparentes, e, portanto, é necessária a presença de profissionais treinados, como os avaliadores, para identificá-las.

Porém, no desenho desta avaliação, chama-se a atenção para outras duas dimensões muito importantes de impacto deixadas pelo animal em sua caminhada: uma acima e outra abaixo do solo. As marcas imprimidas acima do solo correspondem aos arbustos ou galhos que são quebrados ou ao esterco deixado enquanto

o animal faz sua caminhada e correspondem a mudanças causadas pelos projetos no meio ambiente, em políticas ou leis locais/regionais ou em pessoas, grupos e organizações que não estavam envolvidos diretamente nas ações. O impacto deixado abaixo do solo, mais difícil de ser observado, representa as transformações em valores fundamentais, como o senso de lealdade ou solidariedade entre os membros dos grupos participantes e o aumento do capital social.

Em resumo, o modelo de avaliação desenvolvido busca identificar impactos em três níveis:

- I. Familiar/Comunitário (no solo): efeitos observados junto às famílias participantes dos projetos em termos de saúde, bem-estar, educação e meio ambiente, e em que medida esses efeitos transbordaram e chegaram também a atingir as famílias que não participavam diretamente dos projetos.
- II. Para além da comunidade (acima do solo): impactos que ocorrem além da região geográfica onde os projetos são implementados, como nos sistemas educacionais da região, em associações ou cooperativas regionais de agricultores ou mesmo nos órgãos responsáveis pela área de agricultura ou desenvolvimento rural em nível regional ou nacional.

- III. Sistema de valores (abaixo do solo): mudanças observadas nos valores e atitudes adotados pelos membros das comunidades apoiadas pelos projetos em relação ao trabalho em sua propriedade, ao convívio e colaboração com seus vizinhos/comunidade e à maneira como se relacionam com o meio ambiente.

Nestes estudos, os avaliadores lidaram com conclusões sobre impactos que podiam variar significativamente, desde melhorias em habitações e no saneamento básico até mudanças nos valores e na forma como os participantes e a comunidade lidavam com a proteção e recuperação do meio ambiente. A equipe de avaliação teve que utilizar diferentes estratégias para conseguir buscar, de maneira sistemática, as evidências mais relevantes sobre as três principais categorias de impacto:

- I. Como apenas pessoas nativas dos países ou regiões possuíam a fluência linguística suficiente e o conhecimento sobre como chegar a informantes-chave nos governos e outras em organizações relevantes, a equipe de avaliação contratou pesquisadores sociais ligados a universidades ou a organizações do terceiro setor nos diferentes países.

Esses profissionais não podiam ter qualquer conexão com a organização financiadora dos projetos. Eles foram orientados a coletar evidências sobre possíveis impactos ocorridos além das áreas onde os projetos atuavam.

- II. Os avaliadores externos coletaram diretamente evidências sobre os possíveis efeitos das intervenções realizadas nas vidas das famílias e das comunidades atendidas pelos projetos. Eles visitaram os projetos em comunidades rurais ou periurbanas, nas quais entrevistaram famílias em suas próprias casas, selecionadas a partir do princípio de máxima variabilidade, para garantir que elas representassem, da melhor maneira possível, as demais famílias. Realizaram também reuniões abertas a todos os membros das comunidades, reuniram-se com os líderes das comunidades e entrevistaram outras pessoas (não participantes dos projetos) ao acaso, dentro das comunidades ou próximo a elas.
- III. Os avaliadores também identificaram evidências sobre as possíveis mudanças nos valores ou atitudes das famílias e comunidades através das entrevistas já mencionadas e também junto a grupos adicionais.

As evidências de sucesso, necessariamente, precisariam ser relacionadas diretamente a resultados que só poderiam ser alcançados com recursos advindos dos projetos, sejam financeiros, de conhecimento, capacidade de mobilização etc. Há um amplo espectro de exemplos concretos dessas evidências, incluindo a perfuração de novos poços para coleta de água potável, construção de abrigos para animais, melhorias nas casas dos participantes, desenvolvimento de hortas familiares, utilização adequada de medicamentos veterinários, criação/fortalecimento de cooperativas de agricultores, criação de mecanismos de solidariedade a vizinhos passando por necessidade, entre muitos outros. Como já mencionado, outras causas alternativas para explicar as mudanças observadas (livro das causas) eram sistematicamente consideradas no contexto de cada projeto. Se houvesse indicação de que alguma delas havia contribuído especificamente para uma mudança, era preciso estimar, a partir das observações e das entrevistas, a dimensão da contribuição de tal causa alternativa. Contribuiu tanto quanto o projeto? Mais que o projeto? Menos que o projeto? Quanto mais? Quanto menos?

As avaliações foram realizadas em um contexto com algumas limitações importantes. Primeiramente, elas precisavam ser feitas em um interva-

lo de tempo curto. Em segundo lugar, os recursos disponíveis eram limitados. E, finalmente, os projetos já estavam sendo desenvolvidos sem que informações de linha de base tivessem sido coletadas. Para tornar viável o trabalho dos avaliadores, após uma densa análise por parte da equipe de avaliação e de discussões amplas envolvendo diferentes *stakeholders*, foram identificados seis valores básicos que orientavam todo o trabalho realizado pela organização nos diferentes países nos quais ela atua. Foram eles:

1. Atendimento das necessidades básicas para uma vida digna.
2. Gerenciamento e cuidado com os animais.
3. Gestão e proteção do meio ambiente.
4. Educação para um mundo justo e sustentável.
5. Empoderamento das famílias e comunidades.
6. Melhorias nos sistemas e políticas que influenciam diretamente a vida das pessoas.

Esses seis valores possibilitaram que a equipe de avaliação mapeasse os quase 100 critérios e indicadores de avaliação para determinar o impacto dos projetos.

## Comentários finais

---

Finaliza-se sintetizando alguns pontos considerados essenciais para esta discussão:

- a. Dizer que um único método é o padrão ouro é como dizer que um único medicamento é o melhor que existe. Você precisa se perguntar sempre se ele é o melhor para que tipo de problema de saúde.
- b. O verdadeiro padrão ouro é aquele que consegue estabelecer uma argumentação causal que seja consistente e dentro da necessidade de precisão requerida pelo contexto avaliativo, com base em evidências corretas e robustas que, ao mesmo tempo, apoiem e testem criticamente essa argumentação.
- c. Não se quer escolher um único método. De longe, os melhores desenhos

avaliativos ou de pesquisa usam o princípio do multiplismo crítico proposto por Shadish (2004), que significa o emprego de uma combinação de métodos, sendo que os pontos fortes de uns compensam os pontos fracos de outros e vice-versa. Todo método tem suas limitações. Portanto, basear-se em apenas um deles é uma prática inadequada.

- d. Há um último ponto crítico que precisa ser considerado. Não se deve escolher apenas o método ou o conjunto de métodos que parece tecnicamente mais adequado para uma situação. Trabalha-se em situações de vida real, com limitações de recursos. Portanto, é preciso escolher métodos de inferência causal mais custos-efetivos para responder de maneira suficientemente robusta às nossas perguntas avaliativas.

## Referências bibliográficas

---

- CHIANCA, T. A causalidade na avaliação de programas sociais: das divergências ao caminho viável. In: OTERO, M. R. (Org.). **Contexto e prática da avaliação de iniciativas sociais no Brasil: temas atuais**. São Paulo, Peirópolis: Instituto Fonte, 2012.
- COOK, T. et al. Contemporary Thinking About Causation in Evaluation: A dialogue with Tom Cook and Michael Scriven. **American Journal of Evaluation**, v. 31, n. 1, p. 105-117, winter 2010.
- DAVIDSON, E. J. **The nuts and bolts of sound evaluation**. Thousand Oaks, CA: Sage, 2005.
- IMBENS, G.; LEMIEUX, T. (Eds.). The regression discontinuity design: theory and applications. **Journal of Econometrics**, v. 142, n. 2, p. 611-850, february 2008.
- SCRIVEN, M. **Evaluation thesaurus**. Newbury Park, CA: Sage, 1991.
- SCRIVEN, M. A summative evaluation of RCT methodology: & an alternative approach to causal research. **Journal of MultiDisciplinary Evaluation**, v. 5, n. 9, p. 11-24, 2008.
- SCRIVEN, M. The methodology of evaluation. In: TYLER, R., GAGNE, R.; SCRIVEN, M. **Perspectives in curriculum evaluation**. Chicago: Rand McNally. 1967, p. 39-83. (AERA Monography Series on Evaluation, n. 1).
- SCRIVEN, M. **Causes, connections and conditions in history, in philosophical analysis and history**. William Dray, ed., Harper and Row, 1966, p. 238-264.
- SHADISH, W. R. Critical multiplism: a research strategy and its attendant tactics. **New Directions for Program Evaluation**. Special Issue: Program evaluation: a pluralistic enterprise, v. 1993, n. 60, p. 13-57, winter 1993.
- SHADISH, W. R.; COOK, T. D.; Campbell, D. T. **Experimental and quasi-experimental designs for generalized causal inference**. Boston: Houghton Mifflin, 2002.
- ST. CLAIR, T.; COOK, T. D.; Hallberg, K. Examining the internal validity and statistical precision of the comparative interrupted time series design by comparison with a Randomized Experiment. **American Journal of Evaluation**, v. 35, n. 3, p. 311-327, September 2014.
- UNITED NATIONS EVALUATION GROUP (UNEG). **Impact Evaluation in UN Agency Evaluation Systems: guidance on selection, planning and management**. UNEG Guidance Document, 2013.