

MAIS QUE BOAS INTENÇÕES

Técnicas quantitativas e qualitativas na avaliação de impacto de políticas públicas*

Mariana Batista

Universidade Federal de Pernambuco (UFPE), Recife – PE, Brasil. E-mail: mariana.bsilva@gmail.com

Amanda Domingos

Universidade Federal de Pernambuco (UFPE), Recife – PE, Brasil. E-mail: ardlima4@gmail.com

DOI: 10.17666/329414/2017

Médicos usam evidências quando prescrevem tratamentos. Formuladores de políticas deveriam usar também.¹
The Economist

Introdução

Formular políticas públicas é uma constante da atividade governamental. Contudo, o conteúdo dessas políticas pode variar bastante. Quando escolhas devem ser feitas, decisões tomadas e áreas priorizadas, o foco passa a ser a base do julgamento sobre quais políticas devem ser formuladas e quais devem ser interrompidas ou reformuladas. Esse é o

* O título desse artigo é inspirado no livro de Kaplan e Appel (2011), *More than good intentions: how a new economics is helping to solve global poverty*.

Artigo recebido em 12/07/2016

Aprovado em 23/01/2017

raciocínio por trás do que ficou conhecido como “políticas baseadas em evidências” (*evidence-based policy-making*) (Gertler *et al.*, 2011).

Num contexto de limitação da capacidade de intervenção do Estado, tanto por razões orçamentárias quanto por razões relacionadas com os custos de mobilização de informação e tomada de decisão, torna-se importante estabelecer critérios que orientem as escolhas de políticas públicas. Contudo, esses parâmetros podem ser muitos. Adotando uma perspectiva racionalista do processo de formulação de políticas, uma das formas de realizar tais escolhas é por meio de resultados de avaliação.² Para essa perspectiva o processo das políticas públicas pode ser visto como um ciclo, organizado em estágios compostos pela formação da agenda, formulação da política, tomada de decisão, implementação e avaliação (Anderson, 1974).³ Esse último estágio é considerado o momento de renovação do ciclo. Com base nos resultados da avaliação das políticas, subsídios são

identificados para justificar a manutenção, reformulação ou até interrupção das políticas.

A avaliação pode ser de vários tipos e classificada de várias formas diferentes. Quanto ao tempo, a avaliação pode ser *ex-ante* ou *ex-post*. Quanto a quem avalia, a avaliação pode ser feita internamente, pelos próprios gestores do programa, externamente, por avaliadores independentes, ou participante, contando com o *input* dos beneficiários do programa. Quanto ao objeto, a avaliação pode ser concentrada nos processos da política, nos resultados, nos impactos ou ainda na eficiência (Cohen e Franco, 2013; Rossi, Lipsey e Freeman, 2004).

O foco do presente trabalho está na avaliação de impacto. Especificamente, nas diferentes técnicas que podem ser mobilizadas para responder às questões da avaliação. Isto é, o problema foi resolvido? O programa funcionou? Tais questões, no entanto, requerem sofisticação metodológica. Isso porque lidam com problemas diversos, desde a identificação ou o estabelecimento da relação de causalidade entre a política pública e o resultado social de interesse, passando por dificuldades de mensuração de efeitos importantes –, como empoderamento e capital social –, até às diferenças entre os efeitos quantitativos dos programas e a percepção dos beneficiários sobre como foram afetados pela política.

Para oferecer uma visão pluralista, mas não exaustiva, das técnicas disponíveis para avaliar o impacto das políticas públicas, o presente artigo discute abordagens quantitativas e qualitativas através da apresentação das técnicas e da exposição de resultados exemplares de avaliação de políticas utilizando cada uma das abordagens. Argumentamos que a avaliação de impacto de políticas públicas envolve duas perspectivas ou objetivos diferentes. A primeira busca quantificar o impacto das políticas e identificar de forma precisa a mudança nos indicadores de resultado, estabelecendo o programa como causa da mudança observada. A segunda parte da perspectiva dos atores envolvidos, beneficiários e implementadores da política, buscando compreender o efeito das políticas sobre os indivíduos a partir da sua própria percepção. A abordagem quantitativa identifica o impacto e a abordagem qualitativa explica como e por quê.

Na abordagem quantitativa discutimos avaliações com base em experimentos controlados e téc-

nicas que se aproximam desse ideal experimental como o pareamento, o modelo de diferença-em-diferença e a regressão descontínua. Todas essas técnicas se preocupam em “isolar” o efeito causal do programa, lidando com problemas de viés de seleção. Para a abordagem qualitativa apresentamos o uso de técnicas etnográficas, entrevistas em profundidade e grupos focais como formas de captar a perspectiva dos atores na avaliação de políticas, preocupando-nos com relatos e experiências que remetam à qualificação do impacto.

Longe de estabelecer uma divisão entre abordagens quantitativas e qualitativas na avaliação de políticas públicas, este artigo busca mostrar as possibilidades e as limitações de cada técnica, argumentando a favor da abordagem multimétodo como forma de estabelecer avaliações mais completas. O pressuposto é o de que diferentes técnicas respondem a diferentes questões, mas em conjunto buscam o mesmo objetivo, que é a formulação de políticas baseada em evidências.

Para atingir os objetivos estabelecidos, este artigo está organizado da seguinte forma: a próxima seção apresenta os fundamentos da avaliação de impacto. Em seguida discutimos a abordagem quantitativa de avaliação, com foco no experimento controlado como “padrão de ouro” e as técnicas quase experimentais. Na terceira seção apresentamos as abordagens de qualificação do impacto das políticas, com foco em técnicas de observação, entrevista em profundidade e grupos focais. Na quarta, discutimos as possibilidades de combinação de métodos na avaliação de políticas. Por fim, a conclusão. Esperamos que esta discussão contribua para a leitura crítica de resultados de avaliação e também para a utilização das referidas técnicas em novos esforços avaliativos.

Fundamentos da avaliação de impacto

Políticas públicas existem para alterar condições vigentes na sociedade classificadas como negativas. Espera-se que intervenções governamentais apresentem resultados, e para identificar se esses resultados de fato ocorreram é preciso se voltar para a avaliação. A avaliação é considerada o último estágio no ciclo de políticas públicas. Uma vez que um

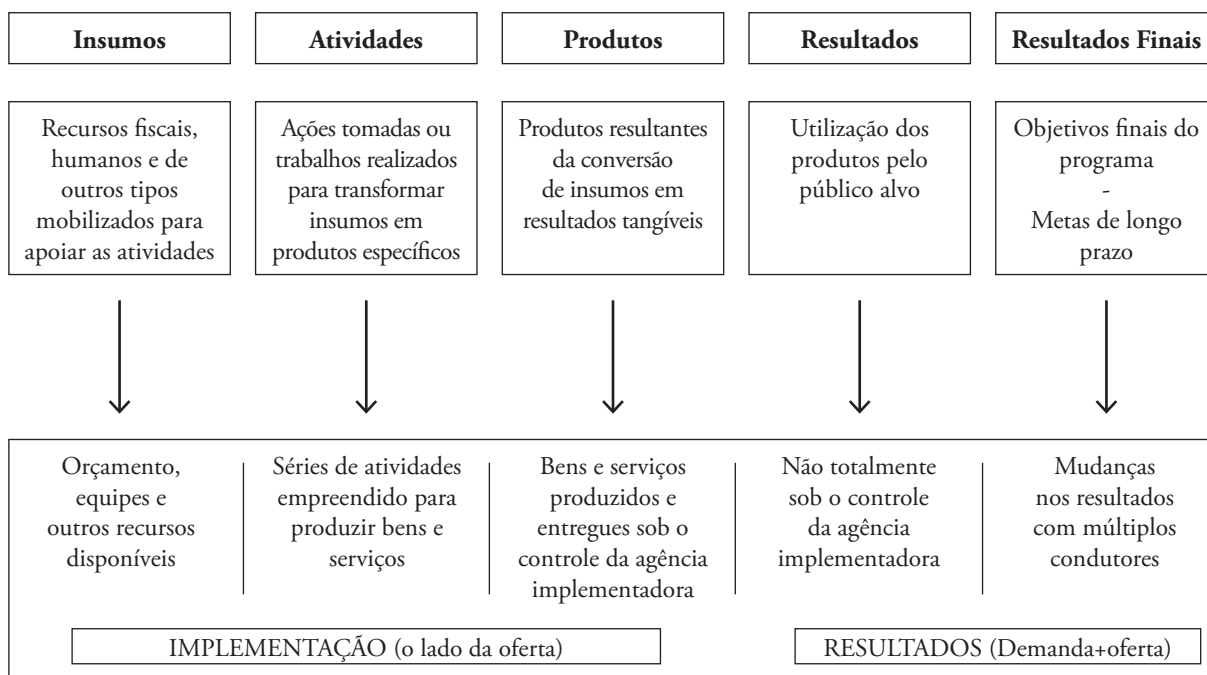
problema social é identificado, uma política pública é formulada e aprovada nas instituições políticas e, após a devida implementação do programa, chega o momento de identificar os resultados de fato do programa (Anderson, 1974).

A avaliação de impacto é um tipo complexo de avaliação, que se diferencia de outros tipo como o monitoramento da implementação e a avaliação de resultados, podendo ser considerada um estágio final do processo avaliativo. De acordo com Rossi, Lipsey e Freeman (2004), a avaliação de processos observa as atividades do programa que de fato foram concretizadas e os serviços que foram entregues ao público-alvo, com foco na cobertura do programa e vieses na entrega. A avaliação de resultados é o acompanhamento da mudança na situação (*outcome*) do público-alvo que o programa busca afetar. O foco dela são as características da população e não as do programa, como na avaliação de implementação, e normalmente são observadas através de metas pré-estabelecidas. A avaliação de impacto, por sua vez, pre-

ocupa-se com a identificação da relação causal entre o programa e o resultado na população. Em suma, a avaliação de impacto tem por base uma relação de causalidade entre política pública e o resultado de interesse. “Colocando de forma simples, uma avaliação de impacto acessa as mudanças no bem-estar dos indivíduos que podem ser atribuídas a um projeto, programa ou política em particular” (Gertler *et al.*, 2011, p. 4).⁴

O raciocínio por trás da avaliação de impacto é o de que a própria política pública pode ser expressa por meio de uma cadeia causal de ações inter-relacionadas que buscam atingir o público-alvo e assim gerar resultados. Dessa forma, para formular a pergunta da avaliação corretamente é preciso especificar quais resultados o programa busca alcançar e através de quais canais/ações. Essa especificação é conhecida como a “teoria da mudança” de um programa ou, simplesmente, “teoria do programa”. Segundo Gertler *et al.* (2011), “teoria da mudança é uma descrição de como se espera que uma inter-

Figura 1
Cadeia de Resultados



Adaptado de Gertler *et al.* (2011).

venção entregue os resultados desejados. Descreve a lógica causal de como e por que um projeto, programa ou política vai alcançar os resultados pretendidos” (*Idem*, p. 22).

A teoria da mudança pode ser modelada como uma cadeia de resultados, de forma a especificar os insumos, atividades, produtos, resultados e resultados finais. A Figura 1 apresenta os elementos de uma cadeia de resultados. Nela pode-se notar que o que é aprovado pelas instituições políticas como a política pública deve primeiro ser traduzido em insumos – recursos financeiros e não financeiros, como pessoal – para em seguida ser concretizada em atividades – ações que transformam recursos em produtos a serem entregues ao público-alvo. Esses produtos podem ser identificados como a concretização da política pública na perspectiva do público-alvo. Trata-se, até aqui, de elementos que representam o lado da implementação da política e estão sob direto controle da agência implementadora. Uma vez que os produtos começam a interagir com o contexto e idiosincrasias dos beneficiários, eles passam a produzir os resultados intermediários e os resultados finais, que correspondem aos objetivos finais do programa.

O ponto-chave da avaliação de impacto é que ela busca identificar se o programa atingiu os seus objetivos finais, sendo que estes são a interação entre as ações do programa e o contexto social e as particularidades dos beneficiários, de forma que não está sob direto controle da agência implementadora.⁵ Ou seja, a avaliação de impacto busca isolar o efeito do programa sobre a mudança no resultado de interesse, excluindo variações causadas por outros fatores, como, por exemplo, flutuações das condições socioeconômicas ou de variação das características dos beneficiários.

Essa preocupação com a identificação de relações de causa e efeito fazem da avaliação de impacto uma atividade metodologicamente desafiadora. Afirmar que o resultado final observado deve-se ao programa é afirmar implicitamente que se o programa não tivesse sido implementado o resultado não seria observado. Para isso seria necessário o chamado *contrafactual*. Esse foco no *contrafactual* na avaliação de impacto de políticas públicas fez surgir um forte programa de pesquisa quantitativa

que busca, através do desenho de pesquisa ou de modelos estatísticos, estimar o que aconteceria se o programa não existisse. Essa é uma das razões que fazem a maioria dos trabalhos sobre avaliação de impacto de políticas públicas ser exclusivamente quantitativo (Adato, 2008).

Contudo, para além da identificação do efeito médio programa, é possível pensar a avaliação de impacto também a fim de entender ou explicar o porquê desse impacto sob o ponto de vista dos atores envolvidos. Essa abordagem busca de alguma forma qualificar o impacto, concentrando-se nas pessoas. Segundo Patton (2002), os avaliadores

Devem estar interessados nas histórias, experiências e percepções dos participantes dos programas para além de saber simplesmente quantos entraram no programa, quantos o completaram e quantos fizeram o quê depois. Achados qualitativos na avaliação iluminam as pessoas por trás dos números e colocam rostos nas estatísticas, não para apelas aos corações, apesar de isso poder acontecer, mas para aprofundar o entendimento (*Idem*, p. 10).

Essa perspectiva, apesar de ser minoritária na avaliação de impacto, vem ganhando força por diversas razões, seja por contribuir para a avaliação quantitativa numa espécie de triangulação, seja por avaliar efeitos do programa sobre indicadores de difícil mensuração quantitativa, seja ainda por ajudar a entender por que em algumas situações o programa não funciona como deveria ou como o programa interage e afeta a cultura local, produzindo resultados heterogêneos.

A avaliação de impacto é um dos momentos mais críticos do ciclo de políticas públicas por se referir às mudanças de fato observadas e sentidas pela população. Por essa razão se torna uma atividade que envolve grandes desafios. Segundo Rossi, Lipsey e Freeman (2004), a avaliação “é um função que os avaliadores devem desempenhar com grande cuidado para assegurar que os achados sejam válidos e devidamente interpretados. Pelas mesmas razões, é uma das tarefas mais difíceis desempenhadas pelos avaliadores e, frequentemente, carregadas politicamente (*Idem*, p. 204).

Nas próximas seções alguns desses desafios serão abordados, apresentando-se sistematicamente as principais técnicas utilizadas para quantificar o impacto de políticas públicas ou para explicar o porquê desses efeitos, concluindo com a perspectiva de integração de métodos como a possibilidade mais completa para a avaliação de intervenções governamentais.

Técnicas quantitativas: quantificando o impacto

Suponha uma política que busca aumentar a renda de trabalhadores através de programas de treinamento. A avaliação de impacto teria como objetivo identificar em quanto a renda dos trabalhadores foi aumentada por causa da sua participação no programa. Para avaliar esse impacto seria necessário observar variação. Isto é, para afirmar que a mudança observada foi causada pela política seria preciso observar o que aconteceria com a renda dos trabalhadores caso a política não existisse. Esse é o chamado *contrafactual*.

De modo intuitivo é possível pensar em dois tipos de comparação para avaliar o impacto de uma política pública (Gertler *et al.*, 2011). A primeira é a comparação entre os indivíduos que receberam a política e os indivíduos que não receberam. No exemplo, compararíamos trabalhadores que fizeram o treinamento com trabalhadores que não o fizeram. A diferença de renda observada entre esses dois grupos depois da implementação da política seria considerado seu impacto. A segunda comparação é entre os mesmos indivíduos antes e depois da política. Isto é, comparar a renda dos trabalhadores antes e depois de participarem do programa de treinamento.

Essas duas desenhos mostram-se intuitivos, mas são considerados equivocados pela literatura de avaliação de impacto (Khandker, Koolwal e Samad, 2010; Gertler *et al.*, 2011; Glennester e Takavarasha, 2013). Isso porque nem o grupo de trabalhadores que não recebeu o treinamento nem o dos trabalhadores antes do programa de treinamento oferecem o *contrafactual*, isto é, a estimativa do que aconteceria caso a política não tivesse

sido implementada. Seriam “falsos *contrafactuals*” (*counterfeit counterfactuals*) porque oferecem uma estimativa do impacto da política pública contaminada por outros fatores (Gertler *et al.*, 2011).

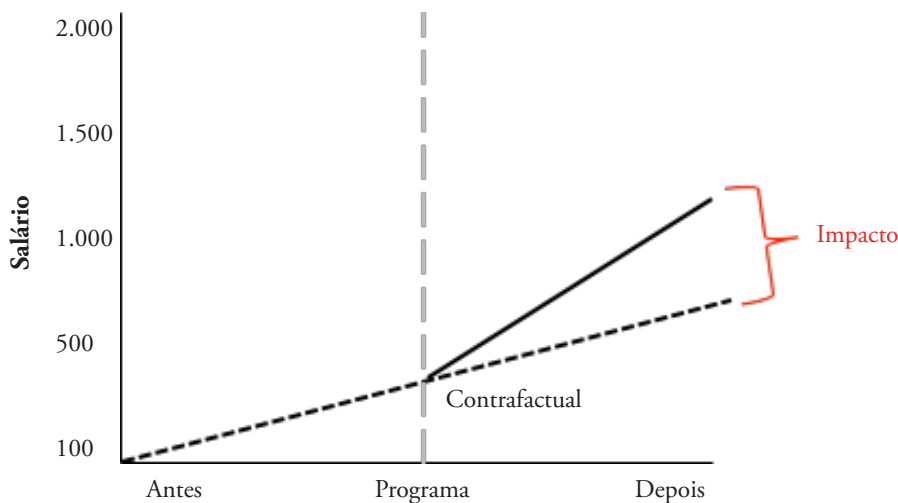
No primeiro caso, da comparação entre os trabalhadores que receberam e os que não receberam o treinamento, a origem da “contaminação” seria o problema de viés de seleção. Isto é, por que alguns trabalhadores receberam o treinamento e outros não receberam? Existe alguma característica sistêmica que diferencie esses dois grupos? No exemplo, poderíamos pensar em características dos próprios indivíduos que fariam com que alguns procurassem participar do programa e outros não. Isso faria com que a diferença na renda entre os dois grupos fosse gerada não só pelo treinamento, mas também por essas características pessoais que fizeram com que os participantes procurassem a política. Dessa forma, a simples comparação entre beneficiários e não beneficiários da política geraria uma estimativa equivocada porque não é capaz de isolar o efeito da política de outros fatores.

No segundo, o da comparação da renda dos trabalhadores antes e depois do treinamento, a estimativa seria “contaminada” por alguma outra coisa que varie no tempo. Por exemplo, a renda dos trabalhadores pode ter aumentado depois do treinamento simplesmente porque a localidade está passando por crescimento econômico e a renda em geral está aumentando. Aqui a mudança na renda seria resultado do treinamento e do desenvolvimento econômico.

O *contrafactual* representa o que aconteceria com esses mesmos trabalhadores que participaram do treinamento caso eles não tivessem participado. Isso quer dizer que estaríamos comparando indivíduos com exatamente as mesmas características e no mesmo contexto socioeconômico. “Para estimar o impacto de um programa, nós precisamos examinar como as pessoas que participaram do programa se sairiam comparadas com como elas se sairiam se elas não tivessem participado do programa” (Glennester e Takavarasha, 2013, p. 24). A Figura 2 mostra a comparação entre esses dois grupos.

A comparação ideal para a identificação do efeito causal seria entre o grupo que recebeu a política e esse mesmo grupo sem ter recebido a política. Contudo, essa situação é claramente inviável.

Figura 2
Identificação do Impacto: O Contrafactual



Elaboração própria.

Porém, há diferentes formas de se aproximar desse ideal hipotético, e um forte programa de pesquisa quantitativa foi desenvolvido para lidar com esse problema. Apresentamos a seguir o experimento controlado e as abordagens quase-experimentais que de uma forma ou de outra estabelecem um grupo de comparação para estimar o efeito causal das políticas públicas.

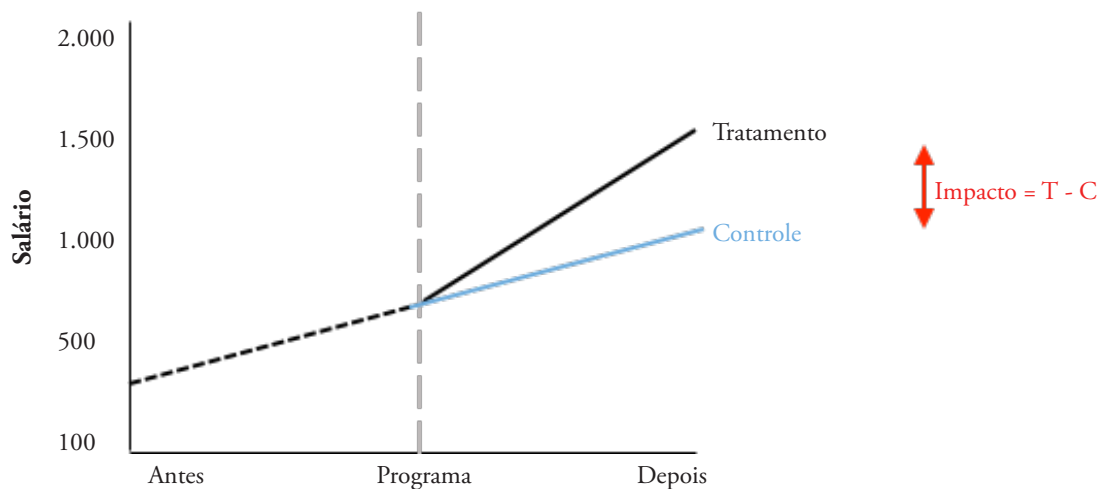
Experimentos controlados: randomização como o “padrão de ouro”

No chamado “padrão de ouro” da avaliação de políticas, a forma de estabelecer o contrafactual seria através do desenho de pesquisa experimental. Dessa forma, os participantes são distribuídos aleatoriamente em dois grupos. Um dos grupos recebe o tratamento (grupo de tratamento), neste caso a política pública, e o outro grupo (de controle) não recebe. A aleatorização estabelece que os participantes tenham a mesma probabilidade de alocação em qualquer um dos dois grupos e que tenham também as mesmas características na média. Os grupos são então equivalentes, significando que não existem diferenças sistemáticas que possam estar correlacionadas com o programa.

A aleatorização na construção dos grupos de tratamento e controle garante que o único fator a distinguir os dois grupos é justamente o recebimento ou não da política pública. Dessa forma, é possível isolar o efeito do programa, excluindo outros fatores que possam influenciar o resultado de interesse e que se confundam com o programa. Dado que os grupos são equivalentes, o efeito do programa seria justamente a diferença entre os dois grupos em termos do resultado de interesse.

No experimento controlado, o desenho da avaliação é estabelecido já no momento da formulação da política, uma vez que para a avaliação alcançar os resultados desejados é necessário ter informações sobre as características dos grupos antes da política (*baseline*) e também que a distribuição entre os grupos seja aleatória, isto é, não obedeça nenhum outro critério sistemático. Isso porque se o recebimento do benefício for associado a uma outra característica sistemática, a diferença observada depois da intervenção foi causada pela política e por essa característica anterior ao programa. Em suma, se a alocação entre os grupos for aleatória e não se observar diferença sistemática entre os grupos, a diferença observada entre os grupos depois da política pode ser atribuída ao programa.

Figura 3
Avaliando o Impacto por Meio do Experimento Controlado



Elaboração própria.

Em termos de técnica estatística, a análise pode ser simplesmente a diferença de média entre o grupo de tratamento e de controle. Pode-se realizar um teste de diferença de média ou uma regressão na qual a variável dependente é o resultado de interesse e a variável independente uma variável binária que assume valor 0 para o grupo de controle e valor 1 para o grupo de intervenção. Claro que em experimentos no campo vários problemas tornam a interpretação dos resultados mais complicada, como *compliance*, *attrition* e *spillovers*. Problemas de *compliance* existem quando uma parcela dos indivíduos a quem são oferecidos o programa não o utilizam e/ou quando indivíduos no grupo de controle recebem o benefício mesmo não tendo sido oferecidos. *Attrition* surge quando indivíduos “saem” do grupo de tratamento. Já problemas de *spillovers* são observados quando o tratamento “transborda” também para o grupo de controle. Esses problemas fazem com que a comparação entre grupo de tratamento e grupo de controle seja dificultada, mas não impossibilitada (Khandker, Koolwal e Samad, 2010).

Outra preocupação central é que a amostra seja grande o suficiente não só para evitar o erro do tipo I (falso positivo, rejeitar a hipótese nula quando não deveria), mas também para evitar o erro do

tipo II (falso negativo, não rejeitar a hipótese nula quando deveria). No primeiro caso o avaliador afirmaria que o programa tem efeito quando na verdade esse efeito é nulo. No segundo caso o avaliador afirmaria que o programa não tem efeito quando na verdade tem. Note-se a implicação normativa e responsabilidade do avaliador, já que o resultado da avaliação diretamente recomenda a manutenção da política ou sua interrupção. Nesse caso, deve-se considerar a importância da *power analysis*, (análise de poder) a qual permite definir o tamanho da amostra requerido para detectar um efeito de determinada magnitude com um nível de confiança especificado. Dessa forma, não somente o desenho, mas também o cálculo específico do tamanho da amostra necessário são fundamentais para uma análise experimental bem-sucedida (Gerber e Green, 2012; Glennerster e Takavarasha, 2013).

Pode parecer difícil pensar a avaliação de políticas públicas com base na distribuição aleatória de benefícios sociais. A randomização pode parecer arbitrária ou antiética. Contudo, alguns governos vêm superando essa resistência inicial ao usar a aleatorização como critério mais justo na alocação de benefícios quando esses são limitados. “Uma das regras mais justas e transparentes na alocação de recursos

escassos para populações igualmente merecedoras acaba por ser dar a todos aqueles que são elegíveis igual oportunidade de participar no programa. Uma forma de fazer isso é simplesmente conduzir uma loteria” (Gertler *et al.*, 2011, p. 49).

Utilizando esse princípio, uma das principais formas de implementar uma avaliação com desenho experimental é a implementação em fases (*phased-in implementation*), na qual os indivíduos elegíveis são divididos em grupos (normalmente geográficos), os quais recebem o tratamento numa ordem aleatória. Assim, a aleatorização entre grupo de tratamento e de controle é garantida e todos os elegíveis receberão a política no final do processo.

Um dos principais experimentos na área de avaliação de políticas públicas foi desenvolvido por Miguel e Kremer (2004), que avaliaram o Primary School Deworming Project (PSDP) no Quênia. O experimento consistiu em randomizar o tratamento contra parasitas em 75 escolas, dividindo-as em três grupos de 25. A intervenção aconteceu durante os anos de 1998 a 2001. No primeiro ano, o Grupo 1 recebeu a intervenção fazendo com que os outros dois fossem grupos de controle. Em 1999 o Grupo 2 recebeu a intervenção, transformando-se em tratamento, e o Grupo III, em controle. Por último, o Grupo III recebeu o tratamento. Os resultados apontaram que as crianças que participaram da política de desparasitação tiveram melhores resultados de saúde e, por conseguinte, maior frequência escolar.

Outro exemplo é o de King *et al.* (2009), uma avaliação do programa Seguro Popular, que foi uma série de reformas implementadas pelo governo mexicano a fim de viabilizar o acesso de pessoas de baixa renda aos serviços de saúde. Os autores utilizam a implementação por estágios para randomizar grupos entre tratamento e controle. Cem *clusters* foram analisados para garantir maior similaridade entre tratamento e controle. Os resultados indicaram que os gastos com saúde diminuíram em 23% em todas as famílias no grupo de tratamento e 55% nas famílias efetivamente tratadas dentro do grupo de tratamento (*compliers*). O programa, no entanto, não apresentou efeito substantivo na utilização de serviços médicos.

Como pode-se notar nos exemplos aqui apresentados, para uma avaliação experimental ser implementada, a preocupação com o desenho da ava-

liação tem de existir no momento da formulação da política, antes da sua implementação. Isso porque a estratégia de inferência sobre o impacto da política se dá com base no desenho (aleatorização) e não em modelos estatísticos. Dessa forma, quando a política pública já foi implementada e sua distribuição já foi realizada (de forma não aleatória), a avaliação por meio do experimento controlado logicamente não é mais possível. A próxima seção discute técnicas estatísticas criadas para lidar com essa situação.

O que fazer quando o experimento não é possível?

Quando o experimento não é possível ou a política já foi implementada não há a separação *ex-ante* entre grupo de tratamento e de controle e, por isso, a identificação do efeito causal da política se torna mais difícil. Quando o que está disponível para o analista são dados observacionais, é possível trabalhar com técnicas estatísticas quase-experimentais para “emular” um grupo de controle ou o contrafactual. Discutiremos três das principais técnicas estatísticas usadas para estabelecer um grupo de controle a partir de dados observacionais: pareamento, diferença-em-diferença e regressão descontínua.⁶

Pareamento

Uma dessas técnicas é o pareamento (*matching*), em que se constrói um grupo de comparação usando características observáveis. Isto é, quando não é possível distribuir de forma aleatória os indivíduos no grupo de tratamento e de controle, o pareamento utiliza as informações sobre os indivíduos como renda, gênero, escolaridade ou outras de acordo com o propósito da política para identificar indivíduos bastante semelhantes nessas características e estabelecer que os dois grupos sejam “balanceados”.

A ideia dessa abordagem é bastante intuitiva: uma vez que o objetivo é isolar o efeito da política pública de outros fatores que possam influenciar o resultado de interesse, a partir de dados observacionais é possível identificar unidades que participaram e que não participaram do programa bastante semelhantes antes da implementação da política, diferindo apenas no recebimento da política, de maneira que as diferenças entre os grupos depois da implementação da política possam ser atribuídas ao programa.

Figura 4
Pareamento Exato

Unidades Tratadas				Unidades não Tratadas			
Idade	Gênero	Meses de trabalho	Diploma ensino médio	Idade	Gênero	Meses de trabalho	Diploma ensino médio
19	1	3	0	24	1	8	1
35	1	12	1	38	0	2	0
41	0	17	1	58	1	7	1
23	1	6	0	21	0	2	1
55	0	21	1	34	1	20	0
27	0	4	1	41	0	17	1
24	1	8	1	46	0	9	0
46	0	3	0	41	0	11	1
33	0	12	1	19	1	3	0
40	1	2	0	27	0	4	0

Adaptado de Gertler *et al.* (2011).

Nessa técnica, o ponto fundamental passa ser a definição do que seria “semelhante”. A ideia inicial seria o pareamento exato. Ou seja, identificar participantes que tenham exatamente os mesmos atributos diferindo apenas no recebimento da política. A Figura 4 apresenta um exemplo de pareamento exato.

A ideia de identificar unidades idênticas tem forte apelo na avaliação por se mostrar bastante apropriada para a identificação do efeito causal. Se as unidades tem exatamente as mesmas características antes da política (*baseline*), sendo o recebimento do tratamento (política pública) o único elemento a diferenciá-las, então qualquer diferença *ex-post* pode ser atribuída à política. Contudo, por mais intuitiva que seja, geralmente é bem difícil encontrar um número grande de observações exatamente iguais em várias dimensões. Sobre o pareamento exato, Gertler *et al.* (2011) argumentam:

Infelizmente, isso é mais fácil de falar do que de fazer. Se a lista de características observáveis relevantes é muito grande, ou se cada característica assumir muitos valores, pode ser muito difícil de identificar pares para cada unidade no grupo de tratamento. Na medida em que se aumenta o número de características ou dimensões com

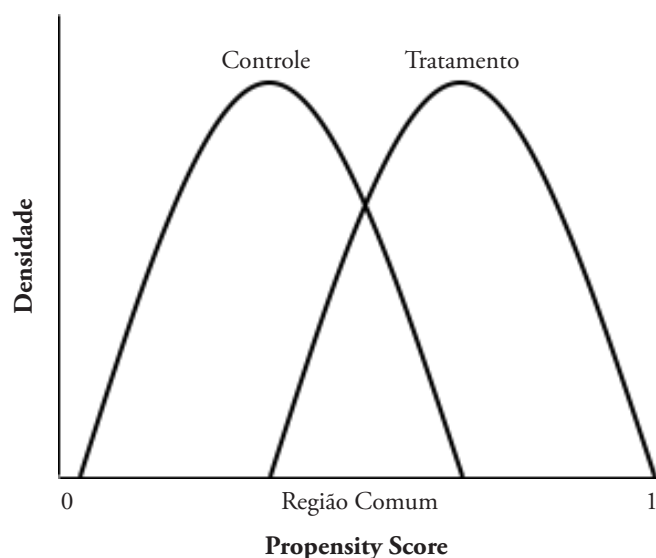
base nas quais se pretende parrear unidades participantes do programa, é possível se deparar com o que é chamado de “maldição da dimensionalidade” (*Idem*, p. 106).

Para lidar com essa limitação, Rosenbaum e Rubin (1983) apresentaram um método de pareamento baseado na “propensão” dos indivíduos de receberem o tratamento. O método ficou conhecido como Propensity Score Matching (PSM) e vem sendo bastante utilizado na avaliação de políticas públicas.

Os indivíduos semelhantes – um participante e outro (ou outros) não participante da política – são pareados de acordo com a “propensão” de participar. O pesquisador não identifica dois indivíduos iguais, mas é possível identificar indivíduos semelhantes o suficiente para fazer a comparação e assim identificar o efeito da política. O efeito do programa é então a diferença entre os grupos.

A utilização do PSM depende de dois pressupostos. Primeiro, fatores não observáveis, como a autoestima do indivíduo, não podem afetar a participação. O segundo pressuposto é o de que exista uma região comum relativamente grande. Ou seja, é preciso ter sobreposição dos valores de propensão, significando que há indivíduos semelhantes partici-

Figura 5
Propensity Score Matching (PSM) e a Região Comum



Elaboração própria.

pando e não participando da política para ser possível fazer a combinação. A Figura 5 apresenta um exemplo de região comum.

Em termos de técnica estatística, primeiramente é realizada uma regressão logística que estima a probabilidade de participar do programa e na qual as variáveis independentes são as características dos indivíduos levantadas como relevantes. Em seguida, os valores preditos para cada indivíduo são estimados e é identificada a região comum, ou a região de sobreposição da propensão entre participantes e não participantes. Os dois grupos são então combinados. Por último, o impacto é identificado a partir da diferença no resultado entre participantes e não participantes combinados.

Um exemplo de utilização do PSM na avaliação é apresentado em Jalan e Ravallion (2003), que analisam o programa Trabajar, do governo argentino, durante a crise econômica de 1997. Os autores examinaram os ganhos de renda dos beneficiários. O PSM foi feito com base em variáveis que resumem as condições socioeconômicas das famílias, como características do domicílio, dados demográficos e histórico educacional. Com base nos escores estimados, os au-

tores separaram os grupos de tratamento e controle. A partir da comparação entre os grupos os autores não encontraram efeitos positivos do programa.

Outro exemplo é a avaliação do Programa Bolsa Família pelo governo brasileiro (MDS, 2007). O estudo estima o impacto do programa sobre gasto de consumo, educação e trabalho infantil, antropometria, saúde e poder de barganha das mulheres. Fora estabelecido um grupo de tratamento, composto por famílias que declararam receber o benefício e dois grupos de comparação: (c1) elegíveis, que não recebem o benefício, mas recebem outros programas; (c2) não elegíveis, que nunca receberam nenhum tipo de benefício. Os resultados indicam que os participantes tiveram melhor desempenho.

Apesar de ter se tornado uma das técnicas mais populares na avaliação de políticas públicas, o PSM vem sendo alvo de críticas. King e Nielsen (2016) argumentam que *propensity scores* não devem ser usados para pareamento porque o pareamento aleatório que o PSM oferece na verdade aumenta o desbalanceamento entre as unidades pareadas. Esse desbalanceamento também aumenta quando cresce o número de variáveis incluídas no procedimento, reduzindo assim a

Figura 6
Diferença-em-Diferença em Prática

	Renda	
	Participantes	Não Participantes
Antes Intervenção	80	90
Depois Intervenção	125	120

Adaptado de Khandker, Koolwal e Samad (2010).

principal contribuição do PSM como ferramenta. Em outras palavras, o pareamento tem como objetivo “podar” as observações de forma que o restante das observações apresentem o melhor balanceamento entre os grupos de tratamento e controle. Como o PSM coloca em primeiro lugar o tamanho da amostra pareada, o método não é explicitamente desenhado para garantir balanceamento, e o processo precisaria ser repetido e ajustado até que se chegasse a um resultado aceitável. O problema com balanceamento surge porque o pareamento pode ser feito facilmente para algumas variáveis, mas pode piorar o balanceamento em outras, gerando a situação na qual o pareamento acaba por aumentar o viés (Iacus, King e Porro, 2012). Como alternativa, Iacus, King e Porro (2011) apresentam o Coarsened Exact Matching (CEM), que retoma a abordagem do pareamento exato, mas com um procedimento de estratificação dos valores das variáveis como maneira de facilitar o pareamento. O CEM inverte o foco e garante que o desbalanceamento entre os grupos pareados não seja maior que um valor escolhido *ex-ante* pelo usuário.

Diferença-em-diferença

A diferença-em-diferença (DD) tem por base a comparação entre indivíduos participantes e não participantes da política antes e depois do programa, daí o seu nome remeter à dupla diferença. Essa técnica parte do pressuposto de que características não observáveis existem e podem afetar a propensão de participar do programa. Com base nisso, a técnica incorpora a heterogeneidade não observável, mas apenas a que não varia no tempo. Dessa forma, tem-se um grupo de tratamento e um grupo de controle (participantes e não participantes), antes e depois da política.

A avaliação é feita simplesmente a partir dessas diferenças, antes e depois para os que receberam a política menos antes e depois para os que não receberam. Essa dupla diferença é justamente o impacto da política. Em termos de técnica estatística é possível calcular as diferenças ou usar uma regressão na qual uma das variáveis independentes é a interação (multiplicação) entre a variável binária participou/não participou e a variável binária antes/depois. O coeficiente da interação é então o efeito da política.

A Figura 6 apresenta um exemplo dessa dupla comparação: tratamento *versus* controle, antes *versus* depois da política.

O DD é uma abordagem para isolar o efeito causal de uma política pública que se baseia nos dois “falsos contrafactuais” apresentados na primeira seção. Na Figura 6, aplicando a equação da dupla diferença tem-se como efeito do programa o aumento de 15 na renda. Contudo, se a avaliação for feita com base na comparação antes/depois esse efeito é de 45, ou, se for feita com base na comparação entre participantes e não participantes, é de 5. Assim, os dois “falsos contrafactuais” se distanciam da estimativa do impacto da política, no primeiro caso superestimando o efeito e no segundo caso subestimando. É importante notar que isso aumenta a discricionariedade do avaliador, de forma que se pode recomendar a manutenção ou interrupção da política, dependendo da técnica de avaliação escolhida.

A avaliação da Lei Maria da Penha (Cerqueira *et al.*, 2015) é um exemplo de avaliação de impacto utilizando o DD. A política tem como objetivo último reduzir os homicídios de mulheres por razões vinculadas ao gênero. Como se trata de uma lei na-

cional, parece impossível num primeiro momento estabelecer um grupo de controle, isto é, um grupo de indivíduos que não tenha sido beneficiado pela política. Isso porque, uma vez que a política foi aprovada, os seus benefícios estão disponíveis para todas as mulheres em todo o território nacional. Para resolver esse problema os autores identificaram o grupo de controle com base no homicídio de homens. Isto é, um grupo de indivíduos que está sujeito às outras causas da violência em geral, como as condições socioeconômicas, e que não é beneficiário da política em questão. O resultado de interesse é a taxa de homicídio, o grupo de tratamento são as mulheres⁷ e o grupo de controle, os homens. O DD é calculado a partir da interação entre ser do grupo de tratamento, depois da política. Os resultados mostram que a lei se mostrou efetiva (*Idem*).

O DD é uma importante ferramenta para a avaliação porque é preciso ter apenas informações antes e depois da política, observar variação quanto ao recebimento, ou seja, um grupo que recebeu a política e um grupo que não recebeu e que os dois grupos tenham a mesma tendência pós-tratamento (Khandker, Koolwal e Samad, 2010). *A priori*, para o uso do DD os dois grupos não precisam ser balanceados nas características pré-tratamento. Exige-se apenas que essa heterogeneidade seja constante no tempo. Em situações em que essas condições pré-tratamento variam no tempo é preciso lidar com o problema para não gerar um viés de variável omitida. Uma forma de lidar com essa situação é combinar o PSM com o DD. Isto é, primeiro parear as unidades e depois calcular a dupla diferença entre os pares (*Idem*).

Regressão descontínua

A última técnica quantitativa a ser apresentada e que vem adquirindo muita atenção recentemente é a regressão descontínua (RDD). Diferentemente do PSM e do DD, a RDD é considerada um “experimento natural”, porque nela a distribuição dos indivíduos entre grupo de tratamento e grupo de controle é feita de forma aleatória, mas essa aleatorização não está sob controle do avaliador (Dunning, 2012).

O foco está nas regras de elegibilidade do programa. Quando os programas são desenhados estabelecem com base em qual critério um indivíduo pode participar do programa. O pressuposto é que

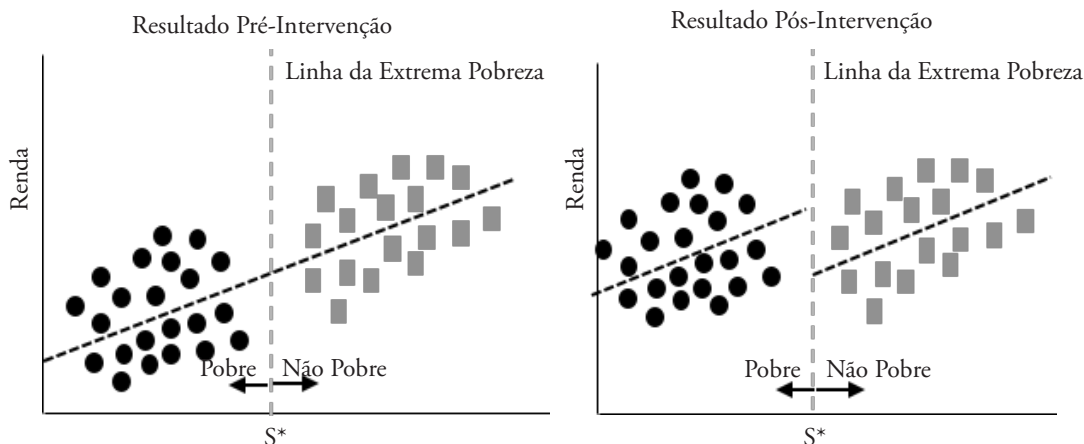
por trás desse critério de elegibilidade existe um índice no qual os indivíduos logo acima e logo abaixo do critério de elegibilidade são bastante similares. O principal atrativo da RDD é que ela permite realizar a avaliação de impacto sem excluir beneficiários: o grupo de comparação são indivíduos logo acima do valor de corte de elegibilidade (*cutoff*). Dessa forma, a técnica estabelece uma comparação válida com base na distribuição “aleatória” dos indivíduos em torno do *cutoff*. O pressuposto é o de que quanto mais próximo desse corte, os indivíduos são bastante semelhantes, sendo o recebimento do tratamento o único fator a diferenciá-los.

“De fato, quando nos aproximamos extremamente perto do escore de corte, as unidades à esquerda e à direita da linha serão tão semelhantes que a comparação será tão boa quanto se tivéssemos escolhido os grupos de tratamento e de controle usando atribuição aleatória do tratamento” (Gertler *et al.*, 2011, p. 91).

Como exemplo, podem-se citar os programas de transferência de renda que estabelecem um determinado valor, acima do qual o indivíduo não é elegível para o programa. No caso da linha de extrema pobreza de 77 reais, indivíduos recebendo até esse valor são elegíveis ao programa, mas indivíduos recebendo 78 reais já não são elegíveis. Contudo, o que diferencia esses indivíduos tão próximos? A lógica é que dentro desse limiar pode-se dizer que os indivíduos foram distribuídos aleatoriamente. Então os participantes e os não participantes podem ser comparados como um grupo de tratamento e um grupo de controle. A Figura 7 apresenta essa comparação.

A RDD é uma ferramenta extremamente poderosa, sendo necessários apenas dois pré-requisitos para o seu uso: 1) um índice de elegibilidade que seja contínuo; 2) um valor de corte claramente definido. Contudo, existem algumas limitações. A primeira diz respeito ao fato das estimativas da RDD serem apenas quanto ao efeito local (em torno do *cutoff*) e não generalizáveis. Isso quer dizer que quanto mais distante do valor de corte, mais heterogêneas serão as observações e por essa razão unidades tratadas e não tratadas não são mais comparáveis. A segunda limitação diz respeito à necessidade de muitas observações em torno do *cutoff*, uma vez que as estimativas são apenas locais.

Figura 7
Descontinuidade na Avaliação de um Programa de Transferência Condicionada de Renda (PTCR)



Elaboração própria.

Em termos de técnica estatística, apesar do nome, a RDD não precisa ser uma regressão. O primeiro passo é avaliar se as observações abaixo e acima do *cutoff* estão “balanceadas”, isto é, não apresentam diferenças sistemáticas além do recebimento do tratamento. O impacto da política seria então o tamanho da descontinuidade (Dunning, 2012).

Buddelmeyer e Skoufias (2004) utilizam a regressão descontínua para analisar a primeira experiência de transferência condicionada de renda: o Programa Progresa, do México, que visa diminuir a pobreza nas áreas rurais e recém urbanizadas. O objetivo é acessar o impacto do programa na dimensão de educação. A descontinuidade utilizada é a elegibilidade da criança para o recebimento do benefício. Os resultados são em relação ao grupo de inelegíveis para a mesma localidade do grupo de tratamento e sugerem que houve um aumento de 5% na presença escolar dos meninos, embora não tenha resultado significativo. A presença escolar das meninas, por sua vez, teve um aumento de 8,9% em relação ao grupo de controle em 1998 e de 9,9% em 1999. No que diz respeito ao trabalho, o programa parece eliminar a maior participação de meninas pobres em atividades de trabalho.

Na avaliação de políticas, a dificuldade de isolar o efeito do programa e poder estabelecer que a mudança no resultado se deve ao programa e não

a outras características acabou gerando essa forte sofisticação metodológica, na qual o desenho experimental ou as técnicas quase-experimentais são as mais recomendadas. O uso de técnicas quantitativas para a avaliação de políticas é a mais difundida já que quantifica o impacto sobre os resultados. Outras características da metodologia quantitativa também contribuem para a sua difusão, como a análise sistemática, a possibilidade de generalização dos resultados, a transparência dos métodos e a possibilidade de replicação dos achados.

Contudo, algumas críticas são feitas à análise estritamente quantitativas. Primeiro, a busca de regularidades e generalização acaba se preocupando com o “efeito médio” ou com o que é comum e não com o que é específico dos indivíduos. A segunda crítica é que o foco nos números acaba por inibir o foco nas pessoas, suas histórias, valores, símbolos e como elas interpretam o mundo. Há ainda a crítica da concentração na explicação e não na interpretação, além da constante exclusão das análises de fatores que não são passíveis de mensuração, mas que ainda assim são relevantes como empoderamento ou capital social (Patton, 2002; Garbarino e Holland, 2009; Schutt, 2015). A próxima seção discute como técnicas qualitativas podem contribuir para o debate, respondendo algumas dessas limitações.

Técnicas qualitativas: qualificando o impacto

Qual é o desenho da avaliação a partir da perspectiva qualitativa e quais as técnicas utilizadas? Partimos do pressuposto de que técnicas quantitativas e qualitativas respondem a diferentes questões na avaliação. A abordagem quantitativa busca identificar a magnitude do impacto e a abordagem qualitativa é mais inclinada para a profundidade. Enquanto uma busca o geral ou o “efeito médio”, a outra busca o que é específico e como o padrão de significação dos atores faz com que a percepção desse efeito seja bem diferente de um indivíduo para outro. Patton (2002) sumariza:

Algumas perguntas se prestam a respostas numéricas; algumas não. Se você quer saber quanto as pessoas pesam, use uma escala. Se você quer saber se são obesas, mensure a gordura corporal em relação à altura e peso e compare os resultados com normas populacionais. Se você quer saber o que o peso significa para as pessoas, como afeta elas, como elas pensam sobre isso e o que elas fazem com relação a isso, você precisa questioná-las, descobrir as suas experiências e ouvir suas histórias. Um entendimento compreensivo e multifacetado do peso na vida das pessoas requer tanto os seus números quanto as suas histórias (*Idem*, pp. 13-14).

A análise qualitativa do efeito das políticas busca primeiramente identificar a qualidade da mudança gerada pela política. Isto é, como as pessoas percebem que sua vida foi melhorada (ou não) pelo programa. O foco está em como os indivíduos percebem a implementação, como foram afetados pelos instrumentos do programa e quais efeitos foram gerados. A análise é feita em termos de valores e significados. Isto é, o que significa para os participantes serem envolvidos numa determinada política, o que significa para eles, por exemplo, aprender a ler ou participar de um programa de treinamento para colocação no mercado de trabalho. O foco vai muito além dos objetivos e metas previstos pelo programa e o interesse está em cada ator como único, com padrão de significação próprio.

“Pesquisa qualitativa é uma abordagem naturalista e interpretativa, preocupada com entender o

significado que as pessoas atribuem aos fenômenos (ações, decisões, crenças, valores, etc.) no âmbito dos seus mundos sociais” (Ritchie e Lewis, 2003, p. 3). Com foco no que é particular, no contexto e no conteúdo da percepção e entendimento dos atores, a abordagem qualitativa oferece elementos para interpretação de como as políticas afetam os indivíduos envolvidos. Contudo, por essas mesmas características essa abordagem envolve a análise de poucos casos, não representativos estatisticamente da população. Como então a pesquisa qualitativa pode ser desenvolvida e contribuir para a avaliação de políticas, a qual geralmente se preocupa com a generalização e com recomendações abrangentes de manutenção ou interrupção de programas governamentais?

A representatividade na perspectiva qualitativa tem a ver com o conteúdo ou mapeamento das visões, experiências e resultados. A ideia é de inclusão e representação simbólica das diferentes posições com relação ao fenômeno em questão (*Idem*). Dificilmente se conseguirá realizar entrevistas em profundidade com uma amostra representativa da população do ponto de vista estatístico, como se faz com *surveys*. Contudo, é possível identificar representantes de diferentes grupos, diferentes características socioeconômicas ou políticas, e a partir dessas diferenças reconstruir ou mapear o fenômeno em questão.

Muitos trabalhos recentes têm partido da perspectiva de que “poucos, se houver algum, estudos avaliativos estariam completos sem incluir alguma informação qualitativa” (Worthen, Sanders e Fitzpatrick, 1997, p. 371). Contudo, buscamos ressaltar aqui que a perspectiva qualitativa na pesquisa avaliativa deve ter um papel em si mesma e não apenas como um “complemento” ao desenho quantitativo.

A vantagem comparativa da pesquisa qualitativa está em analisar e explicar o impacto (Worthen, Sanders e Fitzpatrick, 1997). “Trabalho qualitativo pode ajudar a explicar por que certos resultados são observados na análise quantitativa, e podem ser usados para abrir a caixa-preta do que aconteceu no programa” (Gertler *et al.*, 2011, p. 17). Argumenta-se que a abordagem qualitativa na avaliação é importante por pelo menos quatro razões: 1) parte da perspectiva dos atores para reconstruir a política; 2) identi-

fica impactos diferenciados sobre diferentes grupos e indivíduos; 3) analisa dimensões não diretamente quantificáveis do impacto das políticas; 4) explica por que em algumas situações as políticas atingem os seus objetivos e em outras não.

A perspectiva qualitativa na avaliação vem chamando mais atenção recentemente devido a um renovado interesse em técnicas como entrevista em profundidade e observação participante para entender o processo de implementação da política, além de na interação entre os instrumentos da política e o público-alvo varia de acordo com o contexto e com os padrões de significação de cada ator (Adato, 2008). Aqui analisamos técnicas de entrevista e de observação tendo em vista o seu uso bastante difundido no campo de avaliação.⁸

Observação

“Observações: descrições de campo de atividades, comportamentos, ações, conversas, interações interpessoais, processos organizacionais ou comunitários ou qualquer outro aspecto da experiência humana observável. Os dados consistem em notas de campo, descrições ricas e detalhadas, incluindo o contexto no qual as observações são feitas” (Patton, 2002, p. 4).

A observação tem por base a integração do pesquisador na comunidade pesquisada. O objetivo é a imersão direta na cultura de um grupo. A técnica de coleta da informação (observação), forma de analisar (etnografia) e a teoria subjacente (cultura) normalmente andam em conjunto. A observação pode ser não participante, quando o pesquisador apenas observa, mas não se torna funcional no grupo, mantendo-se externo, ou pode ser participante, quando o pesquisador se torna um membro funcional do grupo, como, por exemplo, um pesquisador que assume a função de professor na avaliação de uma política educacional (Berg, 2001).

O pressuposto da observação como técnica de pesquisa é o de que se o objetivo é entender um povo ou uma comunidade, nada mais apropriado do que conviver com esses indivíduos, observar seus comportamentos, fazer perguntas e ter experiências. Devido a sua origem na antropologia, com o objetivo de compreender comunidades indígenas,

a observação por muito tempo ficou marcada como uma abordagem que somente se adequa a temas específicos desse campo. Contudo, a observação foi redescoberta como abordagem que pode ser utilizada virtualmente para o entendimento de qualquer grupo social, desde comunidades urbanas vistas como caóticas na década de 1930 (Whyte, 1943) até grupos marginalizados no século XXI (Goffman, 2015), passando por qualquer comunidade, grupo, família, beneficiário ou envolvido numa política pública.

De acordo com Patton (2002), “a observação sistemática e rigorosa envolve muito mais do que simplesmente estar presente e olhar ao redor” (*Idem*, p. 5). Esperam-se de um trabalho de observação a análise e a descrição do espaço físico, o desenvolvimento de relações com os observados, o acompanhamento e observação de comportamentos, seguidos de perguntas e esclarecimentos quando o pesquisador não compreender totalmente o que está acontecendo. O tempo aqui é um fator fundamental, tendo em vista que somente observar comportamentos esporádicos num período muito curto pode gerar interpretações parciais. Observações por longos períodos também são recomendadas para diminuir o problema da reflexividade do objeto de estudo. Como o que está sendo observado são seres humanos, espera-se que eles reajam ao observador se comportando de forma diferente do normal. Para isso, recomenda-se que o pesquisador se mantenha na comunidade por períodos prolongados, uma vez que o tempo faz com que o observado naturalize a presença do pesquisador, que acaba se tornando “invisível” (Berg, 2001).

A observação pode ser realizada de maneira menos estruturada na forma de visitas aos locais onde os programas são implementados ou, ainda, para entender como os instrumentos do programa interagem com a cultura local e como sua implementação e seus efeitos variam de acordo com traços da comunidade (Patton, 2002). No primeiro caso a análise é feita com base em descrições densas. No segundo caso essas descrições assumem uma narrativa cultural e são classificadas como etnografias: “etnografia é o trabalho de descrever uma cultura” (Berg, 2001, p. 133).

Na avaliação de políticas a observação é bastante útil principalmente para identificar como a

política pública é percebida no cotidiano do grupo, como ela se adequa ao contexto local, como as pessoas reagem aos seus instrumentos e por quê.

As observações são essenciais para quase todas as avaliações. No mínimo, tais métodos incluem visitas ao local para observar o programa em operação e o uso de habilidades de observação para observar questões contextuais em qualquer interação com as partes interessadas. A observação pode ser usada mais amplamente para aprender mais sobre as operações e resultados do programa, as reações e comportamentos dos participantes, as interações e relações entre as partes interessadas e outros fatores vitais para o estudo (Worthen, Sanders e Fitzpatrick, 1997, p. 375).

Worthen, Sanders e Fitzpatrick (1997) aproveitam para criticar avaliações realizadas por avaliadores que conhecem o programa somente no papel e de fato não têm uma perspectiva de campo de como o programa é entregue ao público-alvo e como essa interação entre os implementadores e os beneficiários pode afetar o impacto dele.

Como exemplo do uso de técnicas de observação na avaliação de políticas podemos citar o trabalho de Adato (2008), que, numa perspectiva multimétodo, faz uma análise de programas de transferência condicionada de renda (PTCRs) na Nicarágua e na Turquia. *Surveys*, etnografia e entrevistas em profundidade são usados para averiguar o impacto do programa sobre saúde, educação e outras variáveis. Três pesquisadores nativos de cada país passaram entre quatro e cinco meses ajudando os habitantes no trabalho, observando e criando confiança para as futuras abordagens de pesquisa. Para a etnografia foi adotado o estudo de caso em nível familiar.

Os resultados mostram diferentes situações com relação aos achados quantitativos. Por exemplo, os resultados quantitativos indicavam que o programa da Nicarágua era extremamente bem focalizado, atendendo o seu público-alvo. Contudo, a percepção dos indivíduos na comunidade é de que o critério de elegibilidade é arbitrário, porque não compreendem como pessoas com renda tão parecida são diferenciadas, algumas sendo classificadas

como elegíveis e outras não. Outro achado é em relação ao suplemento de ferro obrigatório como condicionalidade de saúde. Nesse caso os resultados quantitativos indicavam a distribuição do suplemento, mas o impacto sobre a saúde das crianças não era observado. A observação ajudou a explicar por que o programa não surtiu efeito, uma vez que as mães não administravam o suplemento para as crianças por achar que ele faria mal. Outro resultado interessante é o da pesagem obrigatória das crianças para acompanhar o aumento no peso. A observação revelou que as crianças eram superalimentadas no dia anterior à pesagem como forma de “passar no teste”. Por último, a observação em conjunto com entrevistas revelou um efeito inesperado do programa. O PTCR tem no seu desenho a preocupação com o empoderamento feminino e por essa razão as transferências são feitas diretamente para as mães. Contudo, os homens tinham uma percepção positiva dessa regra por acharem que o cuidado das crianças seria mesmo “obrigação” das mulheres. Dessa forma, esse desenho acabou reforçando uma noção enraizada (Adato, 2008).

Entrevistas em profundidade e grupos focais

“Entrevistas: perguntas abertas e sondagens produzem respostas aprofundadas sobre as experiências, percepções, opiniões, sentimentos e conhecimentos das pessoas. Os dados consistem em citações textuais com contexto suficiente para serem interpretáveis” (Patton, 2002, p. 4).

A entrevista em profundidade é outra fonte de dados comum na avaliação de políticas. Na entrevista a palavra do participante é o foco, sendo as perguntas abertas e flexíveis no sentido de capturar a interpretação do entrevistado, opondo-se assim aos questionários fechados da pesquisa quantitativa. O objetivo é mapear e compreender o mundo dos respondentes. O pressuposto é o de que a percepção que as pessoas têm da realidade se constitui na realidade objetiva delas. O foco está em crenças, atitudes, valores e motivações (Bauer e Gaskell, 2010; Ritchie e Lewis, 2003).

A entrevista pode ser semiestruturada, na qual tópicos predefinidos são estabelecidos. Contudo, entrevistas semiestruturadas geralmente mantêm a

flexibilidade ao permitir inclusão de novos tópicos e permitir a fala do entrevistado. Na entrevista não estruturada não há tópicos predefinidos, garantindo o máximo de liberdade na fala do entrevistado. A narrativa é um tipo especial de entrevista na qual o pesquisador estabelece um fato específico da vida do entrevistado e pede que este narre um determinado processo histórico (Bauer e Gaskell, 2010; Ritchie e Lewis, 2003).

Independentemente do tipo, uma entrevista em profundidade dá centralidade ao entrevistado e não ao questionário como no caso de *surveys*. O objetivo é sempre explorar a percepção dos indivíduos e para isso o entrevistador precisa estabelecer uma relação de confiança, construir um tópico-guia coeso e com progressão lógica plausível de forma a incentivar o entrevistado a expor o seu ponto de vista. Assim, uma entrevista não é algo tão simples ou espontâneo como pode parecer num primeiro momento. “Um bom entrevistar é uma habilidade” (Worthen, Sanders e Fitzpatrick, 1997, p. 380) e, portanto, numa entrevista em profundidade, mais que na aplicação de um questionário, o entrevistador deve estar preparado e ser bem treinado.

Quanto ao número de entrevistados, a entrevista individual busca a profundidade e deve ser utilizada principalmente quando se busca o aprofundamento em determinado tema, o mapeamento da percepção específica do indivíduo ou quando o tema é de alguma forma sensível. A entrevista coletiva tem sua origem na pesquisa de marketing, quando profissionais usam grupos para avaliar um determinado produto. Esse tipo de entrevista evoluiu para o que ficou conhecido como grupo focal.

O grupo focal não é simplesmente a junção de entrevistas individuais; o grupo é diferente da soma de suas partes. O foco está na discussão e na deliberação e não na interação direta entre entrevistador e entrevistados. O entrevistador assume o papel de moderador do grupo, interferindo para fomentar a discussão, limitar participantes dominantes, incentivar participantes relutantes e, por último, para manter o foco da discussão no tema preestabelecido. O uso de grupos focais é especialmente interessante quando se tem por objetivo explorar o espectro de atitudes, opiniões e comportamentos e para observar processos de consenso e divergência

(Bauer e Gaskell, 2010; Ritchie e Lewis, 2003).

Patton (2002) apresenta resultados de entrevistas em profundidade para compreender o impacto diferenciado de um programa de escolarização para jovens e adultos sobre a qualidade de vida dos seus beneficiários. Essas entrevistas foram feitas com grupos de beneficiários que concluíram o programa. A pergunta de interesse foi: que diferença fez nas suas vidas o que vocês estão aprendendo? A seguir reproduzimos algumas das respostas.

- Me ajuda com meu remédio. Agora eu posso ler os frascos e as instruções! Eu estava com medo de dar remédios para as crianças antes, porque eu não tinha certeza;
- Sim, você não sabe o quão embaraçoso é ir às compras e não ser capaz de ler a lista de produtos da esposa;
- Não tenho medo de ler a Bíblia agora na aula bíblica. É realmente importante para mim ser capaz de ler a Bíblia (*Idem*, pp. 16-17).

Os resultados quantitativos indicavam que 77% dos beneficiários estavam “muito satisfeitos” com o programa, embora se note nas falas que por razões bem diferentes. Em comum temos que a ênfase no empoderamento em atividades cotidianas e não em situações relacionadas com a colocação no mercado de trabalho, como seria o objetivo principal do programa. Em suma, o resultado quantitativo indicou a avaliação positiva do programa através de um indicador geral, e as entrevistas em profundidade explicaram como exatamente a vida dessas pessoas foi afetada e por que o programa foi avaliado positivamente.

Outro exemplo de uso de grupos focais e entrevistas individuais na avaliação de políticas é o trabalho de Hunter e Sugiyama (2014) sobre o Programa Bolsa Família (PBF). Três municípios do Nordeste brasileiro foram selecionados para a realização dos grupos: Camaragibe, Jaboatão dos Guararapes e Pau-Brasil. A análise enfocou o Nordeste pela sua histórica associação ao clientelismo e personalismo. Foram realizados onze grupos focais nos três municípios e entrevistas com os gestores e implementadores da política. Os resultados sugerem que o PBF possibilitou que as famílias obtivessem dignidade, por meio da compra de

comida suficiente, roupas e materiais escolares. Afetou também a percepção de independência econômica e que não há constrangimentos por ser beneficiário. De acordo com os grupos focais, houve uma associação entre o PBF e o ex-presidente Luiz Inácio Lula da Silva e o governo federal, impossibilitando a cooperação para o clientelismo por políticos locais. Houve, também, a percepção do voto como mecanismo de *accountability* ao possibilitar a punição aos políticos que se opusessem ao programa. Em suma, portanto, os grupos evidenciaram expresso senso de empoderamento dos beneficiários (*Idem*).

Algumas críticas são comuns à análise qualitativa como a subjetividade e especificidade dos resultados, a ausência de regularidades ou generalização dos resultados, a menor precisão e ainda a dificuldade de replicação. Contudo, oferece um método de avaliação menos intrusivo e preocupado com a visão dos atores. Menos intrusivo justamente porque não impõe uma avaliação exógena, considerando a percepção dos próprios atores envolvidos, oferece ainda a possibilidade de captar aspectos não considerados na análise quantitativa por falta de medidas. Por fim, é responsável pelo conhecimento mais aprofundado, concentrando-se no efeito específico sobre a vida das pessoas, para além do efeito “médio”. Essas contribuições da análise qualitativa em conjunto com os resultados da avaliação quantitativa podem trazer resultados mais robustos e completos sobre o efeito das políticas sobre a vida das pessoas, de forma que o uso de “métodos mistos” se torna uma das principais vias no debate metodológico de avaliação.

Discussão: combinando métodos

“A maioria dos avaliadores agora concorda que nenhum método ou abordagem é sempre apropriado. Em vez disso, o método deve ser selecionado com base na pergunta que se está tentando responder” (Worthen, Sanders e Fitzpatrick, 1997, p. 342). Essa citação reflete a preocupação recente de que a avaliação tem de consistir num projeto integrado, no qual o impacto do programa é identificado e também explicado. Isto é, para fazer mais e melhores políticas precisamos saber quanto foi modificado e o porquê. Precisamos da quantidade

para saber precisamente o efeito do programa, mas precisamos também da qualidade para saber por que tal efeito foi gerado e o que se pode fazer a respeito. Nesse contexto, as limitações de cada técnica específica começaram a ser notadas e a combinação passou a ser vista como o caminho mais frutífero para resolver problemas práticos.

De acordo com Greene *et al.* (2001), “O propósito geral de combinar métodos é proporcionar uma maior redução na incerteza e alcançar uma melhor compreensão dos fenômenos sociais estudados” (*Idem*, p. 30). Há diversas contribuições da abordagem multimétodo para a avaliação de políticas: 1) aumentar a validade e credibilidade das inferências, quando usada para mensurar o mesmo fenômeno buscando a convergência; 2) aumentar a compreensibilidade dos resultados, na medida em que captura diferentes dimensões sob diferentes perspectivas, apresentando uma descrição mais completa do fenômeno estudado; 3) facilitar a descoberta de fatores inesperados; 4) aumentar a diversidade (*Idem*).

Contudo, apesar do relativo consenso em torno da importância da abordagem multimétodo, há diferentes formas de combinação na pesquisa avaliativa. Em algumas situações uma técnica pode claramente sobrepor a outra, enquanto em outras igual importância é atribuída às diferentes formas de analisar o problema em questão. Segundo Caracelli e Greene (1997), os desenhos de pesquisa avaliativa multimétodo podem ser coordenados ou integrados. Quando há coordenação, a combinação de métodos acontece no final do estudo, e um conjunto de evidências é usado para ilustrar, explicar ou refinar os achados de um outro conjunto de evidências. No caso de integração, há pontos de interação entre os métodos ao longo de todo o estudo, desde a formulação dos instrumentos de intervenção, passando pelo desenho da amostra, coleta de dados e interpretação dos resultados finais.

Quando há coordenação, por definição uma técnica se sobrepõe à outra, de forma que o estudo pode ser predominantemente quantitativo ou qualitativo. Quando o estudo é predominantemente quantitativo, a evidência qualitativa é usada para mapear o contexto social, formular os instrumentos de pesquisa e, mais comumente, entender o explicar os resultados quantitativos encontrados. Quando o estudo é predo-

Figura 8
Combinando Métodos

Abordagem predominante	Funcionamento da abordagem predominante	Funcionamento da abordagem complementar
Quantitativa	Questionários estruturados são aplicados a uma amostra aleatória de respondentes e a análise é baseada em testes estatísticos.	Entrevistas, observação e grupos focais são usados para ajudar a formulação do questionário. Pequenas amostras são analisadas para dar maior entendimento às relações estatísticas encontradas na análise quantitativa.
Integrada (quantitativa e qualitativa)	Questionários quantitativos são combinados com formas qualitativas de coleta de dados. Em alguns casos as técnicas qualitativas são usadas para dar contexto e entendimento ao processo em estudo e em outros casos são usadas na mesma unidade de análise como forma de triangulação.	
Qualitativa	Estudos de caso, entrevistas, observação e grupos focais são usados em pequenas amostras de indivíduos.	Questionários são usados para identificar temas e grupos para serem explorados na análise qualitativa. Questionários também são usados para mostrar que a amostra coberta na análise qualitativa é representativa dos grupos na população.

Adaptado de Bamberger (2012).

minantemente qualitativo, a evidência quantitativa é usada para selecionar casos e grupos de interesse, evitando um possível viés na análise. Com a evidência quantitativa é possível mapear a população e identificar casos e grupos de interesse para observação, entrevistas em profundidade e grupos focais. A ideia é mostrar antes ou depois da análise qualitativa que os grupos analisados são representativos dos grupos presentes na população, sendo representatividade aqui sinônimo de variabilidade e não de representatividade estatística. No caso de estudos que combinam métodos através da integração, o uso de técnicas quantitativas e qualitativas é feito em cada estágio da pesquisa, com uma técnica alimentando e ajudando a outra. A Figura 8 sistematiza os tipos de estudos que utilizam métodos mistos na pesquisa avaliativa.

São muitos os trabalhos que utilizam alguma das abordagens acima apresentadas de combinação de métodos, principalmente do primeiro tipo, no qual a abordagem quantitativa é predominante. Por exemplo, Guacituá-Marió, Sians e Wodon (2001) analisam saúde reprodutiva na área rural da Argentina. Nessa avaliação, a análise estatística é combinada a grupos focais com homens e mulheres para melhor

entender o contexto e percepções dos dois grupos. A análise quantitativa é predominante e é usada para identificar o uso de métodos de planejamento familiar e contracepção e o efeito desses hábitos sobre a probabilidade de ter filhos e sobre a atuação no mercado de trabalho. A evidência qualitativa é usada para dar substância e entendimento aos achados quantitativos, mostrando também fatos não evidentes na análise quantitativa, principalmente com relação a temas sensíveis como estupro e incesto.

Clert e Woden (2001) analisam a focalização de programas sociais no Chile, isto é, se os programas estão de fato sendo entregues à população mais pobre. A abordagem quantitativa também é predominante nesse estudo e um *survey* representativo da população nacional é usado para identificar a focalização do programa. Os resultados da análise quantitativa apontam para um grande sucesso na focalização. Combinados ao *survey*, grupos focais são usados para dar significado ao achado quantitativo. Nesse caso, o foco é a percepção dos próprios beneficiários e os achados mostram falhas no programa que não são captadas na análise quantitativa. Os beneficiários demonstram falta de entendimento das regras, o que

por consequência acaba gerando um sentimento de injustiça na distribuição dos benefícios sociais.

Baker (2001) analisa exclusão social no Uruguai e combina técnicas qualitativas e quantitativas, sendo as primeiras dominantes no estudo. O foco do estudo está na relação entre exclusão e pobreza. Métodos quantitativos são a base para a escolha das áreas geográficas onde a análise qualitativa será desenvolvida e para a identificação dos principais grupos em situação de vulnerabilidade. A análise qualitativa é desenvolvida por meio de entrevistas em profundidade e grupos focais e explora as dimensões da exclusão social, bem como as percepções sobre a exclusão em diferentes contextos.

Como exemplo de abordagem integrada podemos citar a avaliação do programa de desenvolvimento comunitário na Índia (Bamberger, 2012). A avaliação utilizou a combinação de métodos de forma sequencial. Primeiro foram selecionadas duzentas comunidade de forma aleatória para designar o tratamento, depois foi realizada uma pesquisa exploratória para entender o contexto da política em termos de direito a terra, participação e redes sociais. Depois, um *survey* foi implementado nas comunidades, seguido de análise em profundidade de cinco projetos que receberam o tratamento e cinco projetos no grupo de controle para observar mudanças na organização das comunidades. O *survey* foi conduzido após dois anos do tratamento e, por último, técnicas quantitativas e qualitativas foram usadas como forma de triangulação para maior entendimento dos resultados alcançados.⁹

Esses exemplos mostram o potencial de combinar métodos quantitativos e qualitativos na avaliação de políticas públicas. A abordagem quantitativa oferece a precisão necessária para o julgamento do funcionamento das políticas, enquanto a abordagem qualitativa oferece o entendimento dos resultados alcançados, mostrando o porquê. Usadas em conjunto, as duas abordagens oferecem elementos para o desenvolvimento de mais e melhores avaliações e julgamentos substantivos, de forma que as políticas alcancem os seus objetivos e de fato melhorem a vida daqueles que são o seu público-alvo.

Conclusão

O presente artigo buscou apresentar as principais técnicas quantitativas e qualitativas na avaliação de impacto de políticas públicas. Iniciamos apresentando os fundamentos da avaliação de impacto e como a busca pelo contrafactual acabou por enfatizar uma tradição de pesquisa quantitativa. O experimento controlado aparece como o “padrão de ouro” para identificação do efeito causal das políticas públicas. Esse desenho, apesar de ser o ideal, muitas vezes não é viável, tanto por razões éticas como por razões pragmáticas. Principalmente quando a avaliação não é pensada no momento da formulação da políticas, o experimento controlado deixa de ser possível. Por essa razão técnicas quase-experimentais foram desenvolvidas para estimar o impacto das políticas. Nesse ponto apresentamos como principais técnicas o pareamento, a diferença-em-diferença e a regressão descontínua.

O foco na identificação do efeito causal fez surgir essa sofisticação metodológica, que do ponto de vista da metodologia quantitativa contribuiu para a precisão das análises. Contudo, a abordagem quantitativa enfoca o “efeito médio” e, por essa razão, acaba deixando de lado a percepção dos atores, suas histórias e como esses efeitos variam de um indivíduo para o outro. Para suprir essa lacuna, técnicas qualitativas vêm sendo cada vez mais usadas. Discutimos o uso de técnicas de observação, de entrevista em profundidade e de grupos focais para a análise de como e por quê as políticas afetam a qualidade de vida dos atores envolvidos.

A conclusão à qual chegamos é que as avaliações são sempre limitadas enquanto estiverem concentradas em apenas uma dessas abordagens. É importante ter uma estimativa precisa do impacto, principalmente para fundamentar decisões quanto à manutenção ou à extinção de programas governamentais. Contudo, essa estimativa está “na média” e não envolve as pessoas por trás dos números. É preciso saber também o porquê dos resultados alcançados. Não pretendemos afirmar a superioridade de uma técnica ou de outra, mas de argumentar a favor da combinação de métodos na avaliação. Combinar métodos aumenta a contribuição de cada um deles, provendo dados mais ricos e um maior poder analítico do que seria possível utilizando os métodos isoladamente.

O objetivo foi apresentar essas técnicas e argumentar a favor da combinação de métodos. Tentamos ser o mais abrangentes e pluralistas possível, dada a limitação de espaço. Não buscamos uma descrição exaustiva desses métodos e tampouco das formas de combiná-los na pesquisa avaliativa. Literalmente, centenas de livros já foram escritos sobre cada um desses pontos. Pretendemos apenas contribuir com a sistematização de parte desse conhecimento para incentivar a produção de mais e melhores avaliações de impacto, especialmente no Brasil.

Notas

- 1 Do original em inglês: “Doctors use evidence when prescribing treatments. Policy-makers should, too” (“In praise of human guinea pigs”. *The Economist*, 12 dez. 2015).
- 2 Enfatizamos que o resultado de avaliações é apenas uma das formas de decidir acerca da manutenção e reformulação de políticas públicas. Em muitas situações o acordo entre os atores envolvidos pode ser um outro critério que não será explorado por não ser o foco do presente trabalho. Em suma, este trabalho apenas discute as técnicas disponíveis para a realização da avaliação uma vez que essa já foi decidida pelos atores políticos.
- 3 Importante destacar, no entanto, que o ciclo de políticas públicas é, muito mais, um tipo ideal do *policy process* do que uma representação fidedigna ao que acontece na realidade (Jann e Wegrich, 2006).
- 4 Nas citações deste artigo, todas as traduções foram feitas pelas autoras, salvo quando indicado outra autoria.
- 5 Apesar de importante instrumento, a utilização da avaliação de impacto pelo governo pode não estar garantida automaticamente, uma vez que as avaliações podem ser um problema para os governantes, executores e gerentes de projetos porque os resultados podem causar constrangimentos públicos (Trevisan e Bellen, 2008). No entanto, sua importância é justificada por ser um mecanismo de melhoria da tomada de decisão, aumento do fluxo de informação e prestação de contas (Ala-Harja e Helgason, 2000).
- 6 As três técnicas foram escolhidas por serem as mais difundidas na avaliação de políticas. Outras técnicas, como o uso de variáveis instrumentais, também são usadas com certa frequência, mas não serão aqui tratadas por limitação de espaço.

- 7 Para uma estimativa ainda mais precisa os autores calculam a taxa de homicídios de mulheres que tenham como local de ocorrência o próprio domicílio, como uma forma de alcançar uma medida mais próxima de homicídio relacionado com questões de gênero.
- 8 Outras técnicas da chamada “nova metodologia qualitativa”, como o Qualitative Comparative Analysis (QCA) ou o Process Tracing, também podem ser utilizadas na análise de políticas, mas ainda com pouca inserção no campo de avaliação. Por essa razão e por limitação de espaço não trataremos aqui dessas técnicas. Para mais informações sobre esses métodos ver Beach e Pedersen (2013), sobre Process Tracing, Ragin e Rihoux (2008) sobre QCA em geral, e Rihoux, Rezsöhazy e Bol (2011) sobre o uso do QCA na análise de políticas públicas.
- 9 Para mais exemplos de combinação de métodos na pesquisa avaliativa, ver www.interaction.org/resources/training/annex-10-case-studies-mm-evaluation-designs-predominant-quant-qual-and-balanced-orientations (consultado em 30/10/2016).

BIBLIOGRAFIA

- ADATO, Michelle. (2008), “Integrating survey and ethnographic methods to evaluate conditional cash transfer programs”. International Food Policy Research Institute. Disponível em core.ac.uk/download/files/153/6337617.pdf, consultado em 30/5/2016.
- ALA-HARJA, Marjukka & HELGASON, Sigurdur. (2000), “Em direção às melhores práticas de avaliação”. *Revista do Serviço Público*, 51 (4): 5-59.
- ANDERSON, James. (1974), *Public policy-making*. Nova York, Holt, Rinehart, and Winston.
- ATTANASIO, Orazio; MEGHIR, Costas & SANTIAGO, Ana. (2012), “Education choices in Mexico: using a structural model and a randomized experiment to evaluate Progress”. *The Review of Economic Studies*, 79 (1): 37-66.
- BAKER, Judy, (2001), “Social exclusion in urban Uruguay”, in E. Gacitua-Mario e Q. Wodon (orgs.), *Measurement and meaning: combining quantitative and qualitative methods for the analysis of poverty and social exclusion in Latin*

- America, Washington, World Bank (Technical Paper n. 518).
- BAMBERGER, Michael. (2012), "Introduction to mixed methods in impact evaluation". *Impact Evaluation Notes*, 3: 1-38.
- BAUER, Martin W. & GASKELL, George. (2010), *Qualitative researching with text, image and sound: a practical handbook for social research*. Londres, Sage.
- BEACH, Derek & PEDERSEN, Rasmus Brun. (2013), *Process-tracing methods: Foundations and guidelines*. Ann Arbor (MI), University of Michigan Press.
- BERG, Laurence. (2001), *Qualitative research methods for the social sciences*. Boston, Pearson.
- BUDELMEYER, Hielke & SKOUFIAS, Emmanuel. (2004), *An evaluation of the performance of regression discontinuity design on PROGRESA*. Washington, World Bank.
- CAIRNEY, Paul. (2011), *Understanding public policy: theories and issues*. Basingstoke, Palgrave Macmillan.
- CERQUEIRA, Daniel; MATOS, Mariana; MARTINS; Ana Paula & PINTO JÚNIOR, Jony. (2015), "Avaliando a efetividade da Lei Maria da Penha". Brasília, Ipea (Texto para Discussão n. 2048). Disponível em www.ipea.gov.br/portal/images/stories/PDFs/TDs/td_2048.pdf, consultado em 5/5/2016.
- CLERT, Carine & WODON, Quentin. (2001), "The targeting of government programs in Chile", in E. Gacitua-Mario e Q. Wodon (orgs.), *Measurement and meaning: Combining quantitative and qualitative methods for the analysis of poverty and social exclusion in Latin America*, Washington, World Bank (Technical Paper n. 518).
- COHEN, Ernesto & FRANCO, Rolando. (2013), *Avaliação de projetos sociais*. Petrópolis (RJ), Vozes.
- DUNNING, Thad. (2012), *Natural experiments in the social sciences: a design-based approach*. Cambridge (MA), Cambridge University Press.
- GARBARINO, Sabine & HOLLAND, Jeremy. (2009), *Quantitative and qualitative methods in impact evaluation and measuring results*. Birmingham, Governance and Social Development Resource Centre/University of Birmingham.
- GERBER, Alan & GREEN, Donald. (2012), *Field experiments: design, analysis, and interpretation*. Nova York, W. W. Norton.
- GERTLER, Paul J. et al. (2011), *Impact evaluation in practice*. Washington, World Bank.
- GLENNERSTER, Rachel & TAKAVARASHA, Kudzai. (2013), *Running randomized evaluations: a practical guide*. Princeton (NJ), Princeton University Press.
- GREENE, Jennifer & CARACELLI, Valerie. (1997), *Advances in mixed-method evaluation: the challenges and benefits of integrating diverse paradigms*. Nova York, Jossey-Bass.
- GREENE, Jennifer; CARACELLI, Valerie & GRAHAM, Wendy. (1989), "Toward a conceptual framework for mixed-method evaluation designs". *Educational Evaluation and Policy Analysis*, 11 (3): 255-274.
- GREENE, Jennifer; BENJAMIN, Lehn & GOODYEAR, Leslie. (2001), "The merits of mixing methods in evaluation". *Evaluation*, 7 (1): 25-44.
- GOFFMAN, Alice. (2015), *On the run: fugitive life in an American city*. Nova York, Picador.
- GACITUA-MARIB, Estanislao, SIAENS, Corinne, & WODON, Quentin. (2001), "Reproductive health in Argentina's poor rural areas", in E. Gacitua-Mario e Q. Wodon (orgs.), *Measurement and meaning: Combining quantitative and qualitative methods for the analysis of poverty and social exclusion in Latin America*, Washington, World Bank (Technical Paper n. 518).
- HUNTER, Wendy & SUGIYAMA, Natasha Borges. (2014), "Transforming subjects into citizens: insights from Brazil's Bolsa Família". *Perspectives on Politics*, 12 (4): 829-845.
- IACUS, Stefano & KING, Gary. (2011), "Causal inference without balance checking: coarsened exact matching". *Political Analysis*, 20: 1-24.
- JALAN, Jyotsna & RAVALLION, Martin. (2011), "Estimating the benefit incidence of an antipoverly program by propensity-score matching". *Journal of Business & Economic Statistics*, 21 (1): 19-30.
- KHANDKER, Shahidur R; KOOLWAL, Gayatri B. & SAMAD, Hussain A. (2010), *Handbook on impact evaluation: quantitative methods and practices*. Washington, World Bank.
- KING, Gary et al. (2009), "Public policy for the

- poor? A randomised assessment of the Mexican universal health insurance programme”. *The Lancet*, 373 (9673): 1447-1454.
- KING, Gary & NIELSEN, Richard. (2015), “Why propensity scores should not be used for matching” [Working paper]. Cambridge (MA), Harvard University. Disponível em gking.harvard.edu/files/gking/files/psnot.pdf, consultado em 8/7/2016.
- MDS – Ministério do Desenvolvimento Social e Combate à Fome. (2007), *Avaliação do impacto do Programa Bolsa Família*. Brasília, MDS.
- MIGUEL, Edward & KREMER, Michael. (2004), “Worms: identifying impacts on education and health in the presence of treatment externalities”. *Econometrica*, 72 (1):159-217.
- PATTON, Michael Quinn. (2002), *Qualitative research and evaluation methods*. Sage (CA), Sage.
- RIHOUX, Benoît & RAGIN, Charles C. (2009), *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*. Sage (CA), Sage.
- RIHOUX, Benoît; REZSÖHAZY, Ilona & BOL, Damien. (2011), “Qualitative comparative analysis (QCA) in public policy analysis: an extensive review”. *German Policy Studies*, 7 (3): 9-82
- RITCHIE, Jane & LEWIS, Jane. (2003), *Qualitative research practice: a guide for social science students and researchers*. Sage (CA), Sage.
- ROSENBAUM, Paul R. & RUBIN, Donald B. (1983), “The central role of the propensity score in observational studies for causal effects”. *Biometrika*, 70 (1): 41-55,
- ROSSI, Peter H; LIPSEY Mark W & FREEMAN, Howard E. (2004), *Evaluation: a systematic approach*. Sage (CA), Sage.
- SCHUTT, Russell (2011). *Investigating the social world: the process and practice of research*. Newbury Park (CA), Pine Forge Press.
- TREVISAN, Andrei Pittol & VAN BELLEN, Hans Michael. (2008), “Avaliação de políticas públicas: uma revisão teórica de um campo em construção”. *Revista de Administração Pública*, 42 (3): 529-550.
- WEGRICH, Kai & JANN, Werner. (2006), “Theories of the policy cycle”, in Frank Fischer, Gerald J. Miller e Mara S. Sidney (orgs.), *Handbook of public policy analysis: theory, politics, and methods*, Boca Raton (FL), CRC Press.
- WHYTE, William Foote. (1943), *Street corner society: the social structure of an Italian slum*. Chicago, University of Chicago Press.
- WORTHEN, Blaine R; FITZPATRICK, Jody L. & SANDERS, James R. (1997), *Program evaluation: alternative approaches and practical guidelines*. Londres, Longman.

**MAIS QUE BOAS INTENÇÕES:
TÉCNICAS QUANTITATIVAS E
QUALITATIVAS NA AVALIAÇÃO
DE IMPACTO DE POLÍTICAS
PÚBLICAS**

Mariana Batista e
Amanda Domingos

Palavras-chave: Avaliação de impacto; Experimentos; Quase-experimentos; Observação; Entrevistas.

Como técnicas quantitativas e qualitativas podem ser mobilizadas para avaliar o impacto de políticas públicas? Este trabalho tem por objetivo apresentar os fundamentos da avaliação de impacto e discutir as principais técnicas utilizadas para quantificar ou para qualificar o impacto das políticas. Para tanto, apresentamos o experimento controlado e técnicas quase-experimentais como ferramentas de identificação do efeito causal das políticas e técnicas de observação, entrevista em profundidade e grupos focais como formas de compreender o seu impacto sobre a qualidade de vida dos beneficiários. As diferentes técnicas são apresentadas com exemplos práticos de políticas públicas. Os resultados indicam que as abordagens quantitativa e qualitativa respondem à questões diferentes na avaliação. Contudo, ambas são fundamentais para a formulação de políticas baseada em evidências.

**MORE THAN GOOD
INTENTIONS: QUANTITATIVE
AND QUALITATIVE TECHNIQS
IN THE IMPACT EVALUATION OF
PUBLIC POLICIES**

Mariana Batista and
Amanda Domingos

Keywords: Impact evaluation; Experiments; Quasi-experiments; Observation; Interviews.

How can quantitative and qualitative techniques be mobilized to assess the impact of public policies? This paper aims to present the basics of impact evaluation and discuss the main techniques used to quantify or qualify the impact of policies. Therefore, we present the controlled experiment and quasi-experimental techniques as tools to identify the causal effect of policies, and observation, in-depth interviews, and focus groups as ways to observe the impact on the beneficiaries' quality of life. Different techniques are presented with practical examples of public policies. The results indicate that quantitative and qualitative approaches answer to different questions in the evaluation. However, both are fundamental to evidence-based policy making.

**DAVANTAGE QUE DE
BONNES INTENTIONS :
TECHNIQUES QUANTITATIVES
ET QUALITATIVES DANS
L'ÉVALUATION DE L'IMPACT DES
POLITIQUES PUBLIQUES**

Mariana Batista et
Amanda Domingos

Mots-clés: Évaluation de l'impact; Expériences; Quasi-expériences; Observation; Entretiens.

De quelle façon les techniques quantitatives et qualitatives peuvent être employées pour évaluer l'impact des politiques publiques ? Cet article a pour objectif de présenter les bases de l'évaluation d'impact et de discuter les principales techniques utilisées pour quantifier ou pour qualifier l'impact de ces politiques. Ainsi, nous présentons l'expérience contrôlée et les techniques quasi-expérimentales tels que les outils d'identification de l'effet causal des politiques et des techniques d'observation, des entretiens approfondis et des groupes de discussion comme des moyens pour comprendre leur impact sur la qualité de vie des bénéficiaires. Les différentes techniques sont présentées avec des exemples pratiques de politiques publiques. Les résultats indiquent que les approches quantitatives et qualitatives répondent aux différentes questions de l'évaluation. Toutes deux sont, cependant, essentielles à la formulation de politiques fondées sur des données probantes.