



estudos  
DEPLAN



**ESTADO DO RIO GRANDE DO SUL**

Secretaria de Planejamento, Governança e Gestão  
Departamento de Planejamento Governamental

# ESTUDOS DEPLAN

*Edição*

*Nº 10/2018*

Avaliação de Impacto de Políticas Públicas  
Conceitos, Metodologias e Experiências

*Junho/2018*



**GOVERNO DO ESTADO DO RIO GRANDE DO SUL**

**Governador:** José Ivo Sartori

**Vice-Governador:** José Paulo Dornelles Cairolí

**SECRETARIA DE PLANEJAMENTO, GOVERNANÇA E GESTÃO**

**Secretário:** Josué de Souza Barbosa

**Secretário Adjunto:** Melissa Guagnini Hoffmann Custódio

**DEPARTAMENTO DE PLANEJAMENTO GOVERNAMENTAL (DEPLAN)**

**Diretor:** Antonio Paulo Cargnin

**Diretora Adjunta:** Carla Giane Soares da Cunha

**EQUIPE EDITORIAL**

Antonio Paulo Cargnin

Juliana Feliciati Hoffmann

**FICHA TÉCNICA:**

Juliana Feliciati Hoffmann e Ana Júlia Possamai (Coordenação e elaboração)

César Stallbaum Conceição, Fernanda Rodrigues Vargas, Gisele da Silva Ferreira, Juliana Feliciati Hoffmann, Sílvia Letícia Lorenzetti, Rayssa Miczewski de Araújo, Vinícius Dias Fantinel (elaboração)

**REVISÃO E TRADUÇÃO:** Marlise Margô Henrich

**CAPA:** Laurie Fofonka Cunha

Estudos DEPLAN / Departamento de Planejamento Governamental - RS.  
Porto Alegre : Secretaria de Planejamento, Governança e Gestão, 2015-

v. : il.

Semestral.

Título especial a cada edição, de acordo com o assunto predominante.

Publicado pela Secretaria de Planejamento, Planejamento, Governança  
e Gestão, 2015-

ISSN 2447-4576

1. Desenvolvimento regional – Periódico – Rio Grande do Sul. I. Rio  
Grande do Sul. Secretaria de Planejamento e Gestão. Departamento de  
Planejamento Governamental.

CDU 332.1(816.5)(05)

Bibliotecário responsável: João Vítor Ditter Wallauer – CRB 10/2016

## AVALIAÇÃO DE IMPACTO DE POLÍTICAS PÚBLICAS: MÉTODOS EXPERIMENTAIS\*

Rafael Dantas<sup>1</sup>

**Resumo:** O texto apresenta o método experimental para avaliação do impacto nas políticas públicas, o qual objetiva identificar a existência (ou não) de uma relação de causa e efeito e, assim, produzir evidência rigorosa sobre a eficácia de um programa ou uma política pública. Será dada ênfase para algumas definições acerca do público-alvo e do grupo controle, sendo este obtido por meio de diferentes técnicas de aleatorização, a fim de assegurar a validade externa e interna do resultado.

**Palavras-chave:** Avaliação do impacto. Políticas públicas. Método experimental.

**Abstract:** This text presents the experimental method to impact evaluation on public policy. Experimental methods aim at identifying the existence (or not) of a cause and effect relationships and then produce criterious evidence about the effectiveness of a program or a public policy. It is presented some definitions about target group e control group. Different randomization techniques to define control group are presented, in order to assure external and internal validation of the evaluation research.

**Keywords:** Impact evaluation. Public policy. Experimental method.

---

\* Texto elaborado a partir da transcrição da fala e da apresentação em slides de palestras proferida no Workshop Avaliação de Impacto de Políticas Públicas, realizado nos dias 29 e 30 de novembro de 2017, no Auditório da Secretaria de Planejamento, Governança e Gestão (SPGG-RS), com o apoio do Banco Mundial.

<sup>1</sup>Coordenador de Campo em Avaliação de Impacto, Banco Mundial.

## **INTRODUÇÃO**

O artigo trata do método experimental para avaliação do impacto das políticas públicas. O método experimental visa identificar a existência (ou não) de uma relação de causa e efeito, produzindo evidência rigorosa sobre a eficácia de um programa. Nas seções seguintes, abordaremos esse tema, destacando algumas definições sobre público-alvo e aleatorização, necessária para a obtenção de uma amostra que produza estimativas que sejam representativas.

## **OBJETIVOS DA POLÍTICA PÚBLICA: COMO MEDIR?**

Os objetivos que uma política pública pretende alcançar são amplos. Por exemplo, pode ser que o objetivo seja melhorar a aprendizagem dos estudantes, aumentar a produtividade das pequenas empresas, interromper a transmissão intergeracional da pobreza, ou engajar o cidadão na política. Mas será que a política pública empregada para alcançar esse objetivo específico funcionou? Quais argumentos podem ser usados para comprovar, ou não, a afirmação de que determinada política foi eficaz?

É importante contestar as afirmações de que a política funciona ou não funciona. Este questionamento trata do viés de seleção que pode haver nos resultados obtidos, uma vez que não necessariamente o programa pode ter incentivado ou estimulado as ações que levaram a tais resultados. Por isso, o objetivo de avaliação de impacto é identificar uma relação de causa e efeito. Ou seja, atribuir exclusivamente ao programa o efeito que se está observando.

## **AValiação DE IMPACTO E O GRUPO CONTRAFCTUAL**

Mas como medir o impacto do programa?

Duas abordagens comumente empregadas são avaliar o que ocorreu antes e depois da intervenção e avaliar o que ocorreu de diferente entre os participantes e os não-participantes. Veremos que essas são duas abordagens ingênuas e que não medem o impacto de uma intervenção.

Na avaliação antes-e-depois, mede-se o resultado da intervenção comparando-o com as métricas obtidas antes da implementação do programa. O problema dessa abordagem é que ela não leva em consideração o efeito do tempo nas métricas avaliadas. Considere uma política de treinamento da mão de obra com o objetivo de aumentar a absorção pelo mercado de trabalho. Algum tempo depois, observamos que os trabalhadores que participaram desse programa estão empregados. Mas poderíamos afirmar que isso se deu por conta do programa que eles participaram? Falhamos em observar que, com o tempo, outras variáveis também

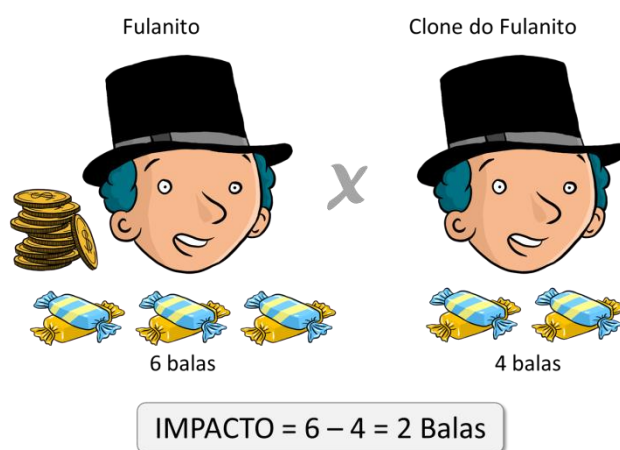
mudam. Por exemplo, pode haver uma melhoria generalizada na economia, aumentando a demanda por trabalho e sendo esse um fator para os resultados que observamos. Não conseguimos assim isolar o efeito da política citada.

Comparar participantes com não-participantes também não mensura o impacto de um programa. Considere a mesma política de treinamento da mão-de-obra e que os candidatos devem se inscrever para participar do programa. Ao final, observamos que a empregabilidade é maior dentre aqueles que se inscreveram e participaram do programa do que aqueles que não se inscreveram e não participaram. No entanto, o ato de se inscrever pode nos dizer muito mais sobre as características dos candidatos que acabaram participando do programa. Estes podem ser os candidatos mais motivados, mais esforçados ou que possuem mais informação sobre os programas disponíveis. E, por conta dessas características, eles teriam, de qualquer jeito, maiores chances de emprego do que aqueles que não se inscreveram. Ou seja, também não conseguimos isolar o impacto da política do impacto das características dos participantes e, portanto, temos uma medida viesada.

A avaliação de impacto, por outro lado, busca medir o resultado que teria ocorrido com o participante caso ele não tivesse participado do programa. Trata-se da chamada análise do contrafactual, por meio da qual se busca comparar o participante nesses dois estados, um em que ele recebe a intervenção e outro em que não recebe, ao mesmo tempo, no mesmo lugar e sob mesmas circunstâncias. Ou seja, é a situação ou evento que não aconteceu, mas poderia ter acontecido.

Por exemplo, suponha que em um experimento uma pessoa receba seis balas para consumo, e que exista, sob as mesmas condições, um clone perfeito dessa pessoa, mas que recebe quatro balas para consumo. A diferença de duas balas é identificada como impacto. A Figura 1 ilustra isso.

Figura 1 – Mundo ideal da avaliação de impacto: comparação entre clones perfeitos

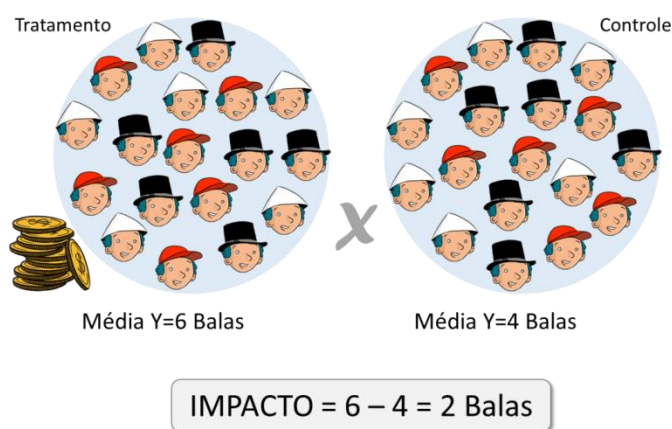


Fonte: Elaboração do autor.

Contudo, medir dois estados para um *mesmo* participante é algo impossível. Para resolver esse problema, podem-se comparar, então, grupos estatisticamente semelhantes, com e sem a intervenção.

Por meio de técnicas estatísticas, especificam-se dois grupos semelhantes entre si: o grupo tratamento que recebe a intervenção (tratamento) e um grupo que não recebe a intervenção (controle). Os dois grupos, tratamento e controle, possuem características iniciais (observáveis e não-observáveis) semelhantes. Entretanto, apenas o grupo tratamento recebe a intervenção, sendo essa a única característica que diferencia um grupo do outro. Assim, a intervenção é a única razão para diferenças observadas após a intervenção. Logo, qualquer diferença observada entre esses dois grupos pode-se atribuir à intervenção (Figura 2).

Figura 2 – Avaliação de impacto na prática: grupos estatisticamente semelhantes



Fonte: Elaboração do autor.

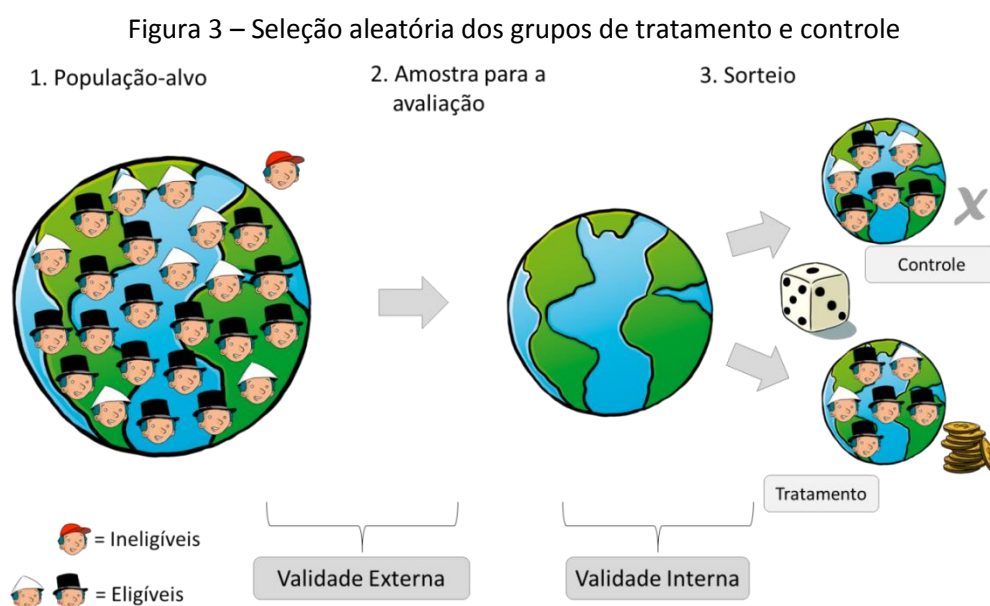
## MÉTODO EXPERIMENTAL: GRUPOS COMPARÁVEIS PARA ASSEGURAR VALIDADE EXTERNA E INTERNA DO RESULTADO

O método experimental é um dos métodos usados para construir os grupos tratamento e controle em uma avaliação de impacto. Nesse método, a definição dos participantes de cada grupo é feita aleatoriamente a partir de uma amostra da população-alvo do programa, composta pelos participantes elegíveis ao programa. A probabilidade de pertencer ao grupo tratamento ou ao grupo controle é igual para todos.

Sendo a amostra aleatória de tamanho suficientemente grande, e o grupo tratamento e controle definidos por sorteio, pode-se concluir que os grupos são comparáveis entre si, sendo a única diferença o tipo de intervenção que um deles receberá. Essa diferença indicará o impacto da política.

Para assegurar que as características, tanto observáveis quanto não-observáveis, estão igualmente distribuídas entre os dois grupos, uma amostra grande deve ser selecionada. Segundo a “lei dos grandes números”, em estatística, uma amostra grande possibilita obter estimativas não viesadas. Ou seja, as estimativas de impacto obtidas são devidas somente à diferença na intervenção.

A partir das etapas de seleção dos participantes em cada grupo, tem-se duas formas de analisar a evidência produzida, seja por validade externa, seja por validade interna. A validade externa é a capacidade de generalização dos resultados para a população-alvo. Ou seja, o impacto observado pode ser estendido à população como um todo. Por sua vez, quanto à validação interna, entende-se que os resultados obtidos evidenciam uma causalidade, pois se estão usando grupos comparáveis. A comparação entre os grupos, então, reflete o impacto naquela população (Figura 3).



Fonte: Elaboração do autor.

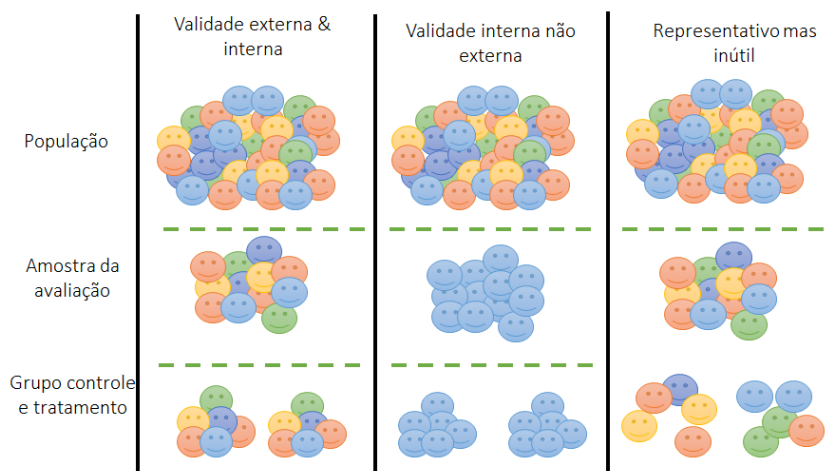
Cuidados com a validade interna e externa são necessários para analisarem-se adequadamente os resultados. A Figura 4 ilustra algumas situações. Na primeira coluna, observa-se uma amostra que é representativa da população-alvo, cujos grupos controle e tratamento foram aleatorizados, sendo representativos da população. Assim sendo, os resultados da análise detêm validade externa e interna.

Na segunda coluna, porém, é apresentado um exemplo em que se tem validade interna, mas não validade externa. A amostra selecionada não é representativa da população, porém produz resultados causais por meio da seleção aleatória dos integrantes que compõem ambos os grupos. Logo, os resultados obtidos valem apenas para essa população em azul.

Na última coluna, por fim, não há validade interna. A comparação está ocorrendo entre pessoas diferentes nos grupos, embora a amostra selecionada seja representativa da

população. Entretanto, o resultado é inútil porque se deseja generalizar um resultado que não é válido para a população como um todo.

Figura 4 – Exemplos de validação



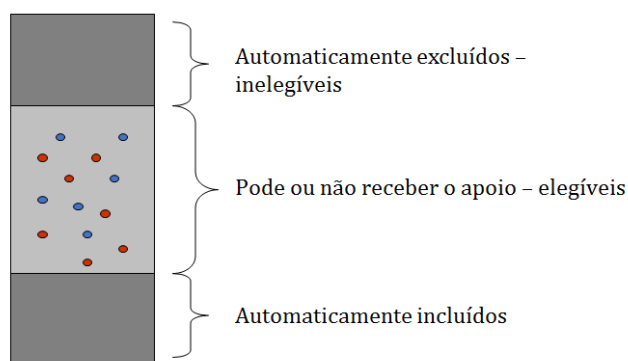
Fonte: Elaboração do autor.

## PÚBLICO-ALVO DA POLÍTICA E PÚBLICO-ALVO DA AVALIAÇÃO DE IMPACTO

Acima se falou de público-alvo e população. Mas qual é o público-alvo do programa? Considere a Figura 5 abaixo. Suponha que exista um programa cujo foco sejam as pequenas e médias empresas. Consequentemente, grandes empresas serão excluídas automaticamente, pois não são elegíveis. Contudo, pode haver um subgrupo que precise de mais assistência, e que, portanto, seja automaticamente incluído no programa. Esse grupo, porém, também será tratado como inelegível para fins de avaliação de impacto.

Por sua vez, um segundo grupo (grupo do meio da Figura 5) pode ou não receber o programa. Esse é um grupo relevante para a avaliação de impacto, sendo considerado um grupo elegível. Dentro desse grupo é que ocorrerá a seleção aleatória do grupo que receberá ou não a intervenção (tratamento e controle). Dessa forma, a avaliação de impacto irá abranger somente esse grupo, e a generalização das conclusões obtidas ocorrerá somente para ele.

Figura 5 – Definição do público-alvo



Fonte: Elaboração do autor.

## OPORTUNIDADES DE ALEATORIZAÇÃO

Compreendido como se compõe o público-alvo de uma avaliação de impacto, cabe discutir como aleatorizá-lo para compor os grupos tratamento e controle. Quando se formulam políticas públicas, há pelo menos quatro fatores que se controla e que são passíveis à aleatorização, que são: o acesso, o *timing*, o encorajamento e a intensidade.

- **Acesso (aleatorização pura):**

Podemos escolher a quem será oferecido o acesso ao programa. As possíveis unidades de aleatorização podem ser: indivíduo ou domicílio, estudantes ou escolas, bairro ou região. Para isso, é preciso ter uma amostra suficientemente grande para conseguir detectar os efeitos. Por exemplo, ao escolher estudantes em vez de escolas, obtém-se uma amostra maior. Isso decorre do fato de que, em uma, existe similaridade entre os estudantes, sendo necessária uma amostra maior de escolas para se detectar o mesmo efeito do que uma amostra de estudantes, independentemente da escola. O agrupamento reduz a efetividade do tamanho da amostra, ainda que reduza os custos de operacionalização e coleta dos dados.

Não obstante, deve ser considerado o risco de contaminação da amostra. Ao se definir os estudantes como nível de aleatorização, por exemplo, por estarem muito próximos um do outro, existe a possibilidade de alguém que esteja recebendo o tratamento influenciar o outro que não está recebendo, ou vice-versa. Nesses casos, pode-se pensar em fazer por escolas, que estão mais separadas umas das outras. Mas a regra geral é tentar aleatorizar na menor unidade possível.

- **Entrada gradual (phase-in design):**

Contexto em que todos devem receber o programa, mas por restrição orçamentária e/ou de logística, não é possível que todos recebam o programa ao mesmo tempo. Portanto, não há como ter um controle puro. Diante disso, pode-se aleatorizar quem receberá primeiro e por último o programa. Nesse cenário, quem recebe primeiro atuará como grupo

tratamento, e quem recebe depois atuará como grupo controle – mesmo que eventualmente venha a receber o programa.

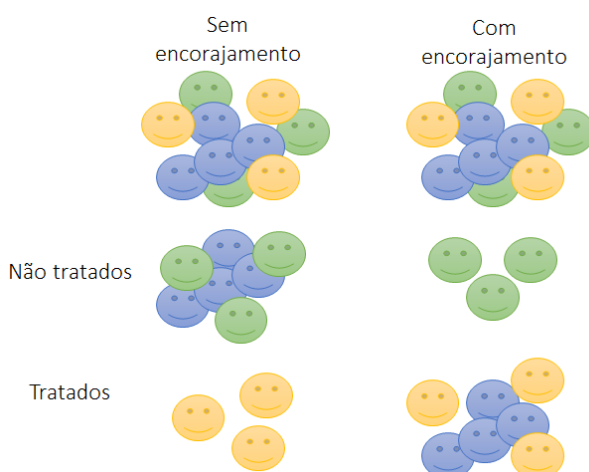
- **Encorajamento:**

Suponha que se lance um programa que ofereça instrumentos financeiros para população carente. O programa tem como objetivo a inclusão financeira e disponibiliza microcrédito, microsseguros, internet *banking*, educação financeira, etc. Pode-se pensar que muitos indivíduos terão interesse no programa. No entanto, pode acontecer o inverso: baixa adesão (*take-up*) – que, na verdade, é um problema bastante frequente nas políticas públicas.

Diante desse cenário, pode-se sugerir um experimento com encorajamento. Dentre os elegíveis, aleatorizar quem receberá um incentivo para ser encorajado a participar (grupo de tratamento) e quem não receberá incentivo (grupo de controle). Esse incentivo deverá afetar apenas a probabilidade da pessoa aceitar ou não o programa, com o objetivo de aumentar adesão ao programa. O grupo relevante para a avaliação de impacto será aquele que responder ao encorajamento e atenderá ao programa.

Portanto, estão sendo considerados três grupos. Têm-se aqueles que decidirão aderir ao programa, os chamados de tratados, e os que não irão aderir, os não tratados. Há ainda aqueles que, independentemente de encorajamento ou não, nunca vão querer ser tratados. A Figura 6 ilustra esse experimento em três tipos de pessoas: os azuis, os amarelos e os verdes. Os verdes são os que não querem ser tratados, independentemente do incentivo. Os amarelos querem sempre ser tratados e sempre irão aderir ao programa, com ou sem incentivo. Já o grupo azul é composto por aqueles que, se não for dado encorajamento, não serão tratados, mas, se for dado, serão tratados. Logo, o grupo relevante para a avaliação de impacto será o grupo azul. Dessa forma, comparamos os integrantes do grupo azul que não receberam o encorajamento, com os que receberam.

Figura 6 – Esboço de encorajamento



Fonte: Elaboração do autor.

- **Variações de tratamento ou de intensidade de tratamento**

Contexto onde todos os elegíveis devem receber o programa, mas, devido à já possuir uma adesão alta, não é oportuno lançar mão do encorajamento. Diante disso, a estratégia é oferecer múltiplos tratamentos.

Por exemplo, um programa de incentivo a pequenas e médias empresas. Pode-se oferecer para um grupo um programa onde será fornecida uma informação sobre quais são as melhores práticas de negócios. Para um segundo grupo, será oferecida uma informação a mais, considerada uma variação nesse tratamento. Dessa forma, serão considerados dois tratamentos e dois grupos, se medirá o impacto da informação mais detalhada.

Outra forma de variação no tratamento é oferecer o mesmo tratamento, mas com intensidades diferentes. Por exemplo, considere um programa de transferência condicional de renda, em que para um determinado grupo seja oferecido R\$ 40,00, enquanto que para outro grupo é oferecido R\$ 60,00, e a outro grupo é ofertada uma dosagem a mais, R\$ 80,00. Com isso, procede-se a uma avaliação de impacto dessas diferentes doses do tratamento.

## **CONSIDERAÇÕES FINAIS**

A contestação acerca das afirmações de que a política funciona ou não funciona leva à avaliação do viés de seleção, que pode ocorrer nos resultados obtidos. Um programa pode ou não ter incentivado, ou estimulado ações que levaram a determinados resultados. A avaliação de impacto auxilia na identificação desses resultados, buscando identificar uma relação de causa e efeito atribuída exclusivamente à intervenção.

As técnicas estatísticas são meios de analisar o impacto da política, a partir da especificação de dois grupos semelhantes entre si. Sendo que o grupo tratamento o que recebe a intervenção, e o grupo controle o que não recebe. Logo, a única característica que diferencia um grupo do outro é a aplicação da intervenção.

O método experimental possibilita definir os grupos, tratamento e controle, aleatoriamente. Para isso, é importante especificar a população-alvo, que é composta pelos participantes elegíveis ao programa, com a finalidade de selecionar uma amostra aleatória de tamanho suficientemente grande. O delineamento a respeito da seleção e validação dos participantes do programa deve ser bem definido para que a construção dos grupos seja representativa e conduza a estimativas não viesadas. Este texto buscou orientar sinteticamente as formas disponíveis para compor esses grupos.

## **REFERÊNCIAS**

GERTLER, Paul J. et al. *Impact Evaluation in Practice*. Washington, DC: Inter-American Development Bank and World Bank, 2016.