

Guia de Avaliação de Impacto Socioambiental

para Utilização em Projetos e Investimentos de Impacto

Insper METRICiS

Núcleo de Medição para Investimentos de Impacto Socioambiental



GUIA GERAL COM FOCO EM VERIFICAÇÃO DE ADICIONALIDADE

SOBRE O INSPER

O Insper é uma instituição independente e sem fins lucrativos, dedicada ao ensino e à pesquisa nas áreas de Administração, Economia, Direito e Engenharia. Tem como missão ser um centro de referência, explorando as complementariedades dessas áreas. Suas atividades de ensino abrangem cursos para várias etapas de uma trajetória profissional: graduação (Administração, Economia e Engenharias), pós-graduação lato e stricto sensu (Certificates, MBAs, programas da área de Direito, Mestrados Profissionais e Doutorado) e Educação Executiva (programas de curta e média duração, e customizados de acordo com as necessidades das empresas). No âmbito da produção de conhecimento, a instituição atua por meio de cátedras e centros de pesquisa que reúnem pesquisadores em estudos e projetos dirigidos a políticas públicas, finanças e gestão. A instituição conta ainda com centros que promovem o empreendedorismo (CEMP) e a educação (Centro de Educação). Tem as certificações de qualidade da Association to Advance Collegiate Schools of Business (AACSB), Association of MBAs (AMBA) e Associação Nacional de MBA (Anamba). O Insper Metricis é um núcleo do Centro de Gestão e Políticas Públicas (CGPP).

SOBRE O INSPER METRICIS

O Núcleo para Medição de Impacto Socioambiental do Insper realiza estudos sobre estratégias organizacionais e práticas de gestão envolvendo projetos com potencial de gerar alto impacto socioambiental. Ênfase especial é dada ao desenvolvimento de ferramentas para planejar, executar e avaliar projetos de impacto realizados por empresas, organizações sem fins lucrativos e governos. Com procedimentos de gestão e avaliação sobre a potencial contribuição de projetos de cunho socioambiental, é possível estimular não somente o crescimento dos investimentos em projetos de impacto, como também impulsionar novas formas de captação e financiamento desses projetos. Além disso, o aprendizado gerado por essas experiências permite o contínuo registro e a disseminação de melhores práticas, por meio de pesquisas acadêmicas, relatórios de políticas públicas, estudos de caso e guias de gestão ligados a projetos de alto impacto.

GUIA DE AVALIAÇÃO DE IMPACTO SOCIOAMBIENTAL PARA UTILIZAÇÃO EM NEGÓCIOS E INVESTIMENE IMPACTO

Uma versão inicial deste documento foi elaborada por Sérgio G. Lazzarini, Leandro S. Pongeluppe, Pui Shen Yoong e Nobuiuki Costa Ito. Esta nova versão recebeu modificações propostas por Lígia Vasconcellos, Sérgio Lazzarini, Carolina Pedrosa Gomes de Melo, Pedro Godoy, José Geraldo Setter Filho, Amanda Arabage, Sergio Firpo e Marina Ribeiro. Revisões anteriores contaram com as contribuições de Mariana Suplicy, Rafael Vivolo, Guilherme Lichand, Amanda Arabage, Carlos Kazunari Takahashi, Sandro Cabral e José Geraldo Setter Filho, além das sugestões de Fernando Carnaúba, Ben Carpenter, Amanda Feldman, Sergio Firpo, Luis Fernando Guedes Pinto, Naercio Menezes Filho, Ricardo Paes de Barros, Brian Trelstad e Mauricio Voivodic. Discussões iniciais para desenvolver este guia evoluíram a partir de um grupo técnico de discussão envolvendo Angélica Rontondaro, Célia Cruz, Franco Veludo, Frederik Kuonen, Simon Locher, Tatiana Fonseca, Rafaella Ziegert e Raquel Costa, com apoio financeiro do Latin America Impact Economy Innovations Fund (Rockefeller Foundation, Fundación Avina e Omidyar Network), em proposta coordenada pelo Instituto de Cidadania Empresarial (ICE), por meio de Célia Cruz e Maria Amélia Sampaio.

São Paulo
Quarta edição,
Janeiro de 2020

INTRODUÇÃO AO GUIA



Há, ainda, muita discussão sobre como monitorar as atividades e medir o impacto de organizações e projetos com foco social ou ambiental. O desafio é ir além do acompanhamento de indicadores, para uma análise dos resultados transformadores causados por um projeto.

Uma forma de se avaliar esses resultados é tentar responder a seguinte pergunta: "O que teria acontecido com os indivíduos ou comunidades-alvo, caso eles *não* tivessem sido beneficiados com o projeto?" Essa pergunta é importante porque, de forma simultânea ao projeto, podem ter ocorrido mudanças externas às intervenções realizadas. Por exemplo, em um projeto buscando apoiar escolas públicas com ferramentas tecnológicas, o gestor do projeto pode erroneamente concluir que essas ferramentas aumentaram o aprendizado dos alunos, quando, na realidade, essa melhoria pode ter sido causada por mudanças em práticas pedagógicas da rede pública.

A chamada abordagem de verificação de **adicionalidade** (*additionality*) busca justamente evitar esse tipo de conclusão errônea. Nessa abordagem, o impacto é definido como a diferença entre o que aconteceu aos indivíduos afetados pelo projeto em relação ao que teria acontecido caso não tivessem recebido as intervenções. Idealmente, se-

ria preciso comparar a *mesma* pessoa ou o *mesmo* grupo recebendo uma dada intervenção. Em alguns casos, isso é possível. Por exemplo, o efeito de usar óculos sobre a capacidade de leitura de pessoas com deficiência visual pode ser medido instantaneamente, a partir de uma simples verificação do quanto a pessoa consegue enxergar com e sem os óculos.

Na maioria dos casos, entretanto, os resultados da intervenção ocorrem ao longo do tempo e podem ser afetados por diversos fatores não diretamente relacionados ao projeto sendo avaliado. Nesses casos, pode-se tentar responder a pergunta "O que teria acontecido com os indivíduos ou comunidades-alvo, caso eles não tivessem sido beneficiados com o projeto?" a partir de informações obtidas de pessoas que, simultaneamente aos beneficiários do projeto, *não* receberam a intervenção.

A ideia é similar ao desenho de um experimento. Por exemplo, no campo da medicina, elege-se um grupo de indivíduos como **grupo tratado**, ou seja, aqueles que receberão o tratamento médico ou o medicamento; e o **grupo de controle**, que não recebe o referido tratamento. Dessa forma, é possível medir como o tratamento afetou a população para além do que poderia ter acontecido naturalmente, sem o tratamento efetuado.

Na abordagem de adicionalidade, o impacto é definido como a diferença entre o que aconteceu aos indivíduos afetados pelo projeto em relação ao que teria acontecido caso não tivessem recebido as intervenções.

Esse cenário alternativo — o que teria acontecido às pessoas sem o tratamento — é chamado de **contrafactual**.

A **Figura 1** exemplifica essa forma de medição, baseada em adicionalidade. Imagine uma empresa que queira investir no desenvolvimento de determinadas comunidades e avaliar o impacto desses investimentos. Para tanto, essa empresa define comunidades que receberão o projeto (o grupo tratado) e comunidades que não serão afetadas pelas intervenções (o grupo de controle). Antes do início do projeto, esses dois grupos encontram-se em pontos distintos de renda, mas têm uma evolução similar em termos do crescimento desse indicador.

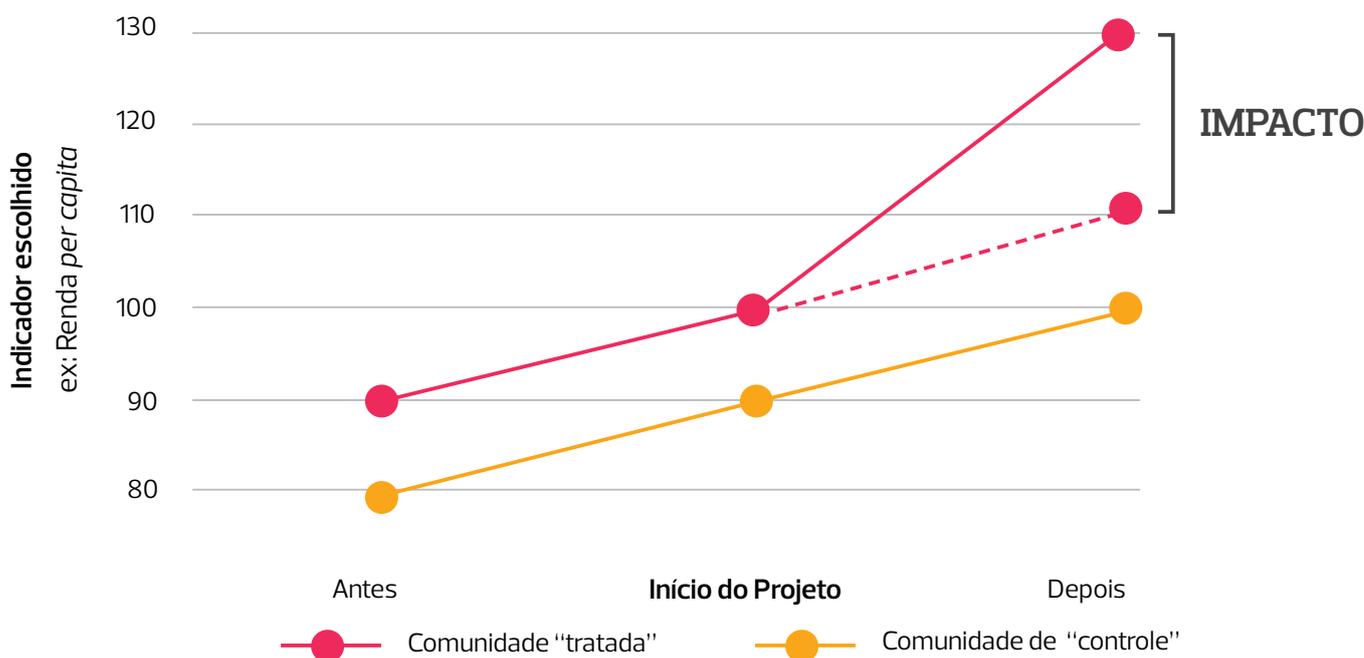
No início do projeto, a empresa realiza novas medições do nível de renda dos indivíduos dos grupos de controle e tratado. Nesse momento, observa-se que o nível médio de renda na comunidade de controle é de \$ 90, e na comunidade a ser tratada (que receberá o investimento), de \$100.

Após a medição, a empresa então inicia o projeto com a comunidade tratada. Após um ano do início do projeto, a empresa realiza novamente a medição do nível de renda individual nas duas comunidades,

tratada e de controle. Um analista inexperiente avaliaria que o impacto do projeto é de \$30 por indivíduo, dada a elevação da renda *per capita* de \$100 para \$130 nos indivíduos tratados. Contudo, com a utilização da metodologia por verificação de adicionalidade, compara-se a evolução do grupo tratado frente ao que provavelmente teria acontecido sem o projeto. Esse cenário contrafactual, nesse caso, é propiciado pela análise do que ocorreu no grupo de controle. Observa-se que na comunidade que não recebeu o investimento houve um aumento de renda de \$90 para \$100, por outras causas que já vinham afetando a localidade (como, por exemplo, melhorias naturais nas condições de vida ou programas de distribuição de renda do governo).

Dessa forma, ao considerarmos não apenas a evolução do grupo tratado ao longo do tempo, mas também a tendência natural que esse grupo seguiria (evidenciada pelo grupo de controle), é possível medir o real impacto do programa. No presente caso, percebe-se que o impacto gerado pelo programa elevou em \$20 a renda de cada indivíduo na comunidade. Esse ganho adicional é simplesmente a diferença entre a evolução verificada na comunidade tratada (\$30) e a evolução verificada na comunidade controle (\$10).

Figura 1: Medição de impacto por verificação de adicionalidade



Seguindo essa abordagem de adicionalidade, o presente guia se baseia em dois pilares essenciais:

1. Indicadores com elevada relevância para aferir melhorias geradas pelo projeto, isto é, resultados positivos esperados na população-alvo.
2. Avaliação com o que poderia ter acontecido a essas populações na ausência do investimento (por exemplo, por meio de comparação com grupos de controle similares que não foram contemplados pelo projeto).

Note-se que monitorar indicadores não é a mesma coisa que avaliar o seu impacto, quando o impacto for definido em termos de adicionalidade. Gestores de projetos sociais usualmente coletam indicadores dos seus grupos tratados ao longo do tempo (o item 1 da definição cima), o que é certamente uma boa prática para acompanhar a sua evolução e o seu desempenho. No exemplo anterior, de um projeto para aumentar o desenvolvimento de comunidades, pode-se acompanhar diversos indicadores, como nível de renda, educação, saúde e vários outros. Contudo, para se verificar efetivamente se o projeto *causou* essas mudanças, seria preciso examinar o cenário contrafactual envolvendo o que provavelmente aconteceria às comunidades sem o projeto (o item 2 da definição acima).

Dessa forma, o *Guia para Avaliação de Impacto Socioambiental para Utilização em Investimentos e Negócios de Impacto* tem a finalidade de prover um ferramental prático que possibilite a investidores, organizações sem fins lucrativos, empresas e governos avaliar o impacto que causam os projetos em que investem. O *Guia* também serve aos próprios gestores dos projetos, sejam eles

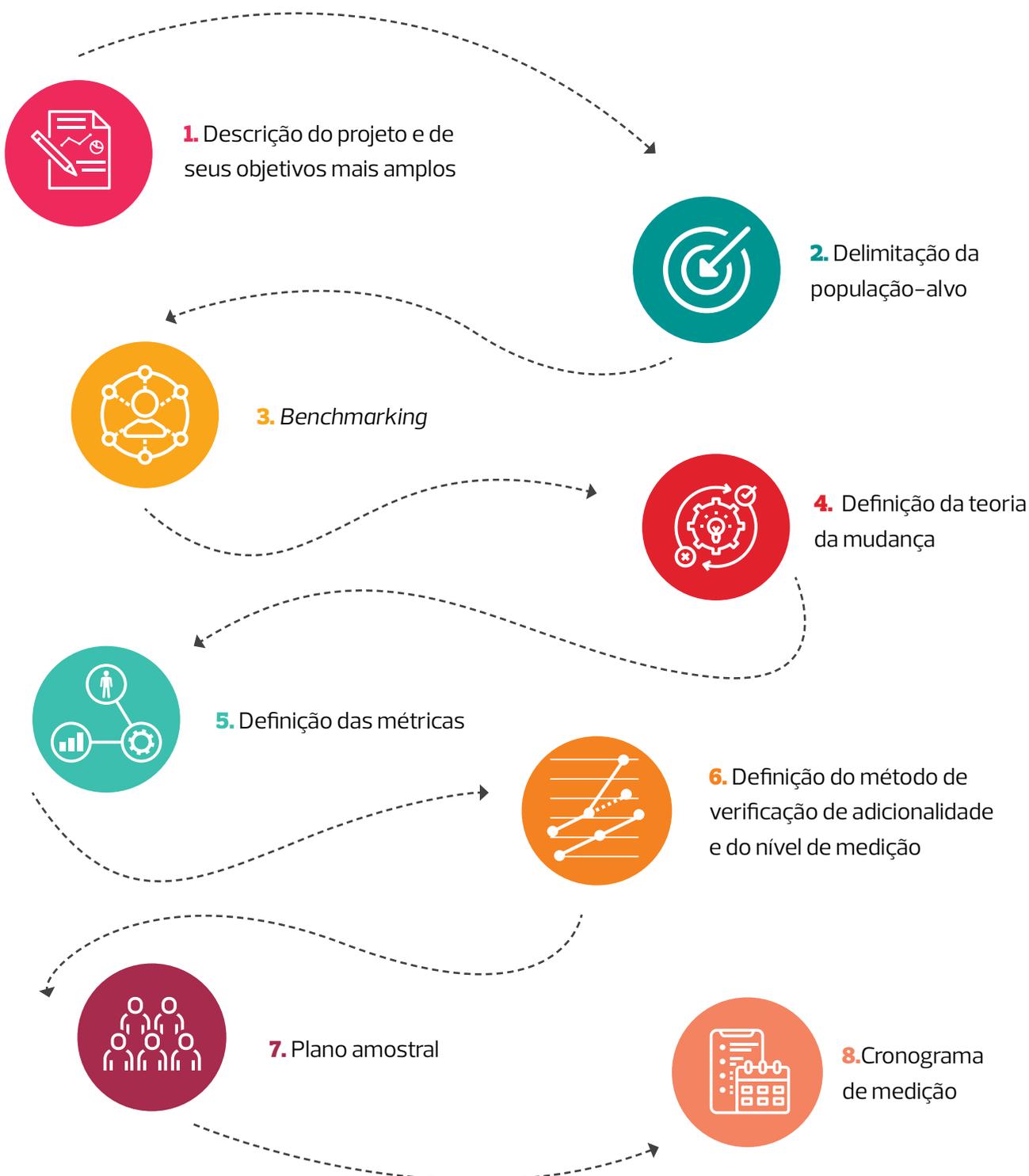
de entidades privadas, públicas ou de terceiro setor, pois os resultados dos projetos são também informação essencial para decisões de gestão e planejamento.

O *Guia* propõe uma sequência de passos que vão desde a definição dos objetivos do projeto e da sua população-alvo, até o processo de escolha de métricas e formas alternativas para cômputo de adicionalidade. Nesse sentido, é uma ferramenta útil para a prática de gestão de impacto. Seguindo a abordagem proposta pelo *Impact Management Project (IMP)*¹, por exemplo, o *Guia* auxilia mais especificamente na dimensão de **contribuição**, que analisa quanto os resultados da intervenção foram diferentes do que teria acontecido sem o projeto. Porém, alinhado ao proposto pelo IMP, o *Guia* também traz uma discussão sobre **quem** são os potenciais beneficiários (definição da população-alvo), **quais** os resultados mais relevantes para eles (definição da teoria de mudança e escolha de métricas) e **quanto** se espera de melhoria com base nesses resultados esperados (a partir da análise de estudos prévios e das próprias medições a serem realizadas ao longo do tempo). ■

1. Detalhes sobre o IMP podem ser obtidos em <https://impactmanagementproject.com/>. Acesso em: 3 jul. 2019.

ELABORAÇÃO DO PLANO DE MEDIÇÃO

O plano de medição é essencial para desenhar, operacionalizar, executar e controlar a medição. Com a finalidade de implementar uma medição robusta e acurada, é importante seguir as seguintes etapas:





1. Descrição do projeto e de seus objetivos mais amplos

A definição de estratégia para se estimar o impacto da intervenção inicia-se com a descrição da intervenção a ser avaliada. Nessa etapa, devem ser conhecidos a missão e os objetivos da organização com o projeto, as ações planejadas e, caso tenham sido implementadas, quando elas foram implementadas e em que contexto elas ocorreram.

Em gestão de impacto, tem sido comum ligar os objetivos gerais do projeto aos **Objetivos de Desenvolvimento Sustentável (ODSs)** das Nações Unidas, que têm sido crescentemente usados para identificar questões-chave para aumentar o bem-estar de populações ao redor do mundo². É possível que um mesmo projeto esteja relacionado a vários ODSs. Por exemplo, considere um projeto para aumentar o acesso de comunidades a água limpa e infraestrutu-

tura de saneamento. Esse projeto é diretamente relacionado ao ODS 6: "Assegurar a disponibilidade e gestão sustentável da água e saneamento para todos". Porém, ao mesmo tempo, o projeto pode afetar positivamente a saúde da população-alvo e a infraestrutura sustentável de cidades, logo, ligando-se também aos ODS 3 (saúde e bem estar) e 11 (cidades e comunidades sustentáveis) respectivamente.

Muitas vezes, apesar de as ações serem muito bem definidas por seus idealizadores, os objetivos são pouco discutidos e detalhados. Um aprofundamento da discussão dos objetivos pode também ajudar a entender quais deles são diretamente influenciados pelo programa, e quais, apesar de desejáveis e relacionados com as ações implementadas, estão fora do campo de ação do programa.

2. UNITED NATIONS. Sustainable development goals. 2019. Disponível em: <http://www.un.org/sustainabledevelopment/sustainable-development-goals/>. Acesso em: 30 jun. 2019.

3. Mais detalhes sobre caracterização da população-alvo podem ser vistos em: IMPACT MANAGEMENT PROJECT. Who. 2019. Disponível em: <https://impactmanagementproject.com/impact-management/what-is-impact/who/>. Acesso em: 03 jul. 2019.



2. Delimitação da população-alvo

Um passo crucial para a definição de um plano de medição é a delimitação da sua **população-alvo**. Para a delimitação objetiva dessa população é necessário, em primeiro lugar, identificar todos os atores relevantes que possam ser afetados pelo projeto ou influenciar seus resultados (os chamados *stakeholders*). Esses atores podem incluir financiadores, provedores de serviço, organizações de suporte e, especialmente, os beneficiários das intervenções. É importante, nessa etapa, verificar se suas prioridades e objetivos estão em linha com as ações planejadas para o projeto.

As organizações geralmente têm recursos limitados, e a escolha de uma população-alvo também é uma oportunidade de centrar esforços em quem realmente precisa ou de dar ênfase a segmentos nos quais a intervenção pode ser mais efetiva.

Por exemplo, um projeto pode ter como objetivo mais amplo promover melhorias de saúde em populações carentes, por meio

do acesso dos cidadãos de baixa renda a medicamentos e formas de tratamento efetivas de doenças crônicas, a um menor custo. Mas quais são essas populações carentes e quais indivíduos poderão ter prioridade, especialmente em uma condição de escassez de recursos?

Fatores possíveis de serem considerados no processo de focalização incluem³:

- **Geografia:** a delimitação regional do projeto.
- **Características demográficas:** idade, gênero, raça e outras.
- **Características socioeconômicas:** renda, escolaridade, ocupação e outras.
- **Necessidades especiais:** deficiências físicas e cognitivas e outras vulnerabilidades relevantes.
- **Susceptibilidade a fatores ambientais:** clima, riscos de desastres e outros.
- **Condições iniciais:** como o segmento focalizado se encontra em termos de variáveis de interesse do projeto.

A escolha de uma população-alvo também é uma oportunidade de centrar esforços em quem realmente precisa ou de dar ênfase a segmentos nos quais a intervenção pode ser mais efetiva.

O **BOX 1** apresenta um exemplo de focalização da população-alvo considerando alguns dos fatores acima indicados.

BOX 1. Exemplo de focalização da população-alvo em um contexto de ações pré-natal com mulheres grávidas.

Um programa governamental em parceria com uma organização sem fins lucrativos busca realizar consultas de saúde pré-natal, para melhorar os indicadores de saúde materna e dos recém-nascidos. O governo definiu um conjunto de municípios totalizando 6.024 mulheres grávidas.

Estudos anteriores indicam que a eficácia desses programas é maior no caso de mulheres vulneráveis com idade entre 14 e 35 anos (*característica demográfica*), com até sete anos de escolaridade (*característica socioeconômica*) e em primeira gravidez (*condição inicial*).

O Diagrama de Venn ao lado indica a quantidade de mulheres com cada uma dessas características e o total de casos na sua intersecção: 732 mulheres. Essas mulheres poderiam ser consideradas como prioritárias para as intervenções propostas, ainda que outros segmentos também possam ser incluídos no programa, caso existam recursos suficientes.



3. Benchmarking

A expansão dos projetos de impacto depende de um esforço de aprendizado contínuo e consolidação de erros e acertos de projetos similares, bem como do uso de conhecimento prévio de medição gerado por pesquisas e estudos. Dessa forma, a análise de casos de sucesso prévios no setor de impacto (*benchmarking*) é essencial: para auxiliar a construção da teoria da mudança (Etapa 4) e a definição de indicadores (Etapa 5), é preciso levantar o que já foi feito ou estudado na atividade e população-alvo do projeto.

Fontes de informação incluem estudos acadêmicos, publicações de organizações internacionais, relatórios de outros patrocinadores de projetos na mesma atividade, entre outras. Há ainda diversas plataformas na internet agregando pesquisas prévias de impacto em diversas

áreas, com informações detalhadas sobre diversas possíveis intervenções e sua efetividade⁴. Em estudos de impacto, tem sido também cada vez mais comum verificar os resultados de *meta-análises*: estudos que tentam agregar os resultados de diversas avaliações prévias sobre um determinado tema.

Como exemplo, imagine um gestor interessado em ajudar pessoas desempregadas a encontrarem um bom posto de trabalho (logo, ligando-se ao ODS 8, enfatizando trabalho decente e crescimento econômico). Uma meta-análise publicada em 2018 analisou 207 estudos sobre programas de ativação de emprego, que incluem treinamento e intermediação para o trabalho, assim como subsídios para empresas⁵. O estudo mostra que os resultados observados após um ano da interven-

4. Por exemplo, a biblioteca de avaliações de impacto do J-PAL (Abdul Latif Jameel Poverty Action Lab, <https://www.povertyactionlab.org/evaluations>); a ferramenta de busca de intervenções em educação da Education Endowment Foundation (<https://educationendowmentfoundation.org.uk/resources/teaching-learning-toolkit>); e os estudos compilados pela Cochrane Library em saúde (<http://www.cochrane.org/evidence>). Acesso em: 4 jul. 2019.

5. CARD, D. et al. What works? A meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, v. 16, n. 3, p. 894–931, jun. 2018.

ção são mais altos que no curto prazo (até um ano). Comparando-se os tipos de intervenção, vê-se que programas com ênfase em treinamento têm resultados mais efetivos do que outros programas que fazem apenas intermediação com empresas.

A etapa de *benchmarking* também permite quantificar o **tamanho do efeito esperado** para o projeto. Por exemplo, na meta-análise citada anteriormente, os autores concluem que esses programas de treinamento podem aumentar a probabilidade de emprego no longo prazo em até 8,7 pontos percentuais. Esse efeito é maior do que o resultado observado após apenas um ano (1,6 ponto percentual), o

que, por si só, já sugere que devem ser observados resultados em mais longo prazo. Caso a revisão indique efeitos pouco relevantes, os gestores podem também tentar mudar ou complementar suas intervenções previstas.

É possível, ainda, que o efeito do projeto seja maior ou menor, dependendo de características heterogêneas da população, como discutido na Etapa 2. Ainda no exemplo de programas de emprego, o gestor pode querer colocar foco em jovens recém-egressos do Ensino Médio e com menor renda. O *benchmarking* deveria, então, localizar estudos especificamente direcionados a essa população-alvo.

A expansão dos projetos de impacto depende de um **esforço de aprendizado contínuo e consolidação de erros e acertos de projetos similares, bem como do uso de conhecimento prévio de medição gerado por pesquisas e estudos.**



4. Definição da teoria da mudança

Antes de aprofundar a questão de como medir o impacto do projeto, primeiramente é necessário definir o que medir. Nessa etapa, o projeto/programa/negócio a ser avaliado deve apresentar sua **teoria da mudança**: uma forma clara e lógica de articular a conexão entre as atividades realizadas e os resultados socioambientais pretendidos.

A **Figura 2** apresenta as cinco etapas de uma teoria da mudança, cujo mapeamento é essencial em um plano de medição: **insumos** e **atividades**, que geram **produtos** ofertados à

população-alvo (**outputs**); os quais, por sua vez, conduzem a **resultados (outcomes)** ligados tanto às atividades principais do projeto, quanto a transformações mais amplas para a sociedade. Essa etapa se beneficia diretamente das etapas anteriores do plano: a definição dos objetivos do projeto e da população-alvo direciona a identificação de resultados mais específicos almejados, enquanto a análise de *benchmarking* permite identificar quais potenciais insumos e atividades podem ser mais efetivos para gerar esses resultados.

6. MCLAUGHLIN, J. A.; JORDAN, G. B. Using logic models. In: NEWCOMER, K. E.; HATRY, H. P.; WHOLEY, J. S. (Eds.). **Handbook of practical program evaluation**. San Francisco: Jossey-BASS, 2013, p. 7-32. HEHENBERGER, L.; HARLING, A. M.; SCHOLTEN, P. **A practical guide to measuring and managing impact**. European Venture Philanthropy Association, 2013, p. 9.

Figura 2: Teoria da mudança aplicada a investimentos e negócios de impacto



Fonte: Adaptado pelos autores com base em McLaughlin e Jordan (2004)⁶

Em geral, gestores de projetos socioambientais tendem a reportar produtos em vez de resultados. Para ilustrar essa distinção, considere uma empresa cuja atividade seja fornecer ferramentas tecnológicas para melhorar os resultados acadêmicos de alunos da rede pública de ensino, utilizando insumos diversos, como as próprias ferramentas fornecidas, com o apoio de equipes de suporte para coordenar a sua aplicação. Os produtos dessa atividade seriam o número de acessos à ferramenta na escola ou o número de sessões de interação com alunos. Observe, entretanto, que nenhum desses produtos indica necessariamente que os alunos melhoraram como resultado da intervenção. É preciso, portanto, analisar quais são os resultados esperados na população-alvo, isto é, melhorias que efetivamente podem causar impacto positivo na vida dos beneficiários ou na sociedade de forma mais ampla.

Nesse caso, um resultado ligado à atividade do projeto seria a melhoria do aprendizado dos alunos, medida por algum teste de conhecimento. Um resultado social mais amplo, por sua vez, seria o aumento esperado de bem-estar e até mesmo ganhos de renda futura em função das novas habilidades adquiridas pelos alunos. Nesse processo, é importante descrever o **mecanismo causal** que irá determinar as mudanças. Por exemplo, as ferramentas podem incluir jogos e vídeos, ajudando a motivar os alunos para o aprendizado e facilitando a compreensão e retenção de conhecimento.

A teoria de mudança permite também definir um painel de indicadores de acompanhamento dos beneficiários do projeto. No exemplo acima, a empresa pode acompanhar as horas alocadas pela equipe de suporte ao projeto (insumos), as ações que ela realiza com os alunos (atividades), o número de acessos à ferramenta (produtos) e o aprendizado dos alunos (resultados), dentre outros indicadores. Porém, sob uma ótica de adicionalidade, a teoria de mudança, em si, não avalia o impacto do projeto. O impacto, como discutido anteriormente e detalhado

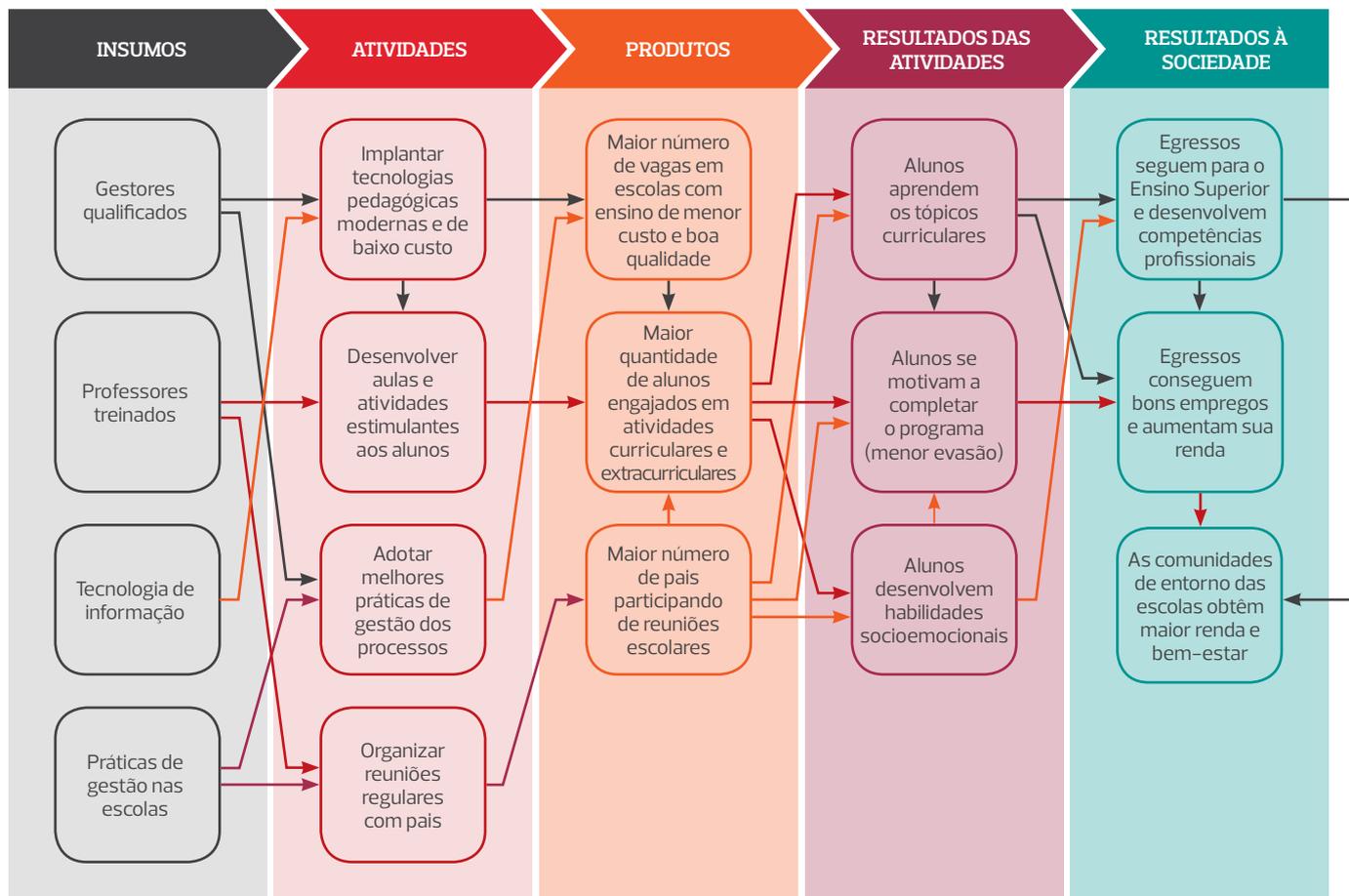
na Etapa 6, envolve resultados que ocorreram devido ao projeto, não a outros fatores externos que possam ter afetado a população-alvo. Monitorar um painel de indicadores da população-alvo é importante, mas não suficiente para avaliar impacto.

O **BOX 2** apresenta um outro exemplo de aplicação de teoria de mudança na área educacional, incluindo um conjunto mais amplo de insumos e intervenções, e com uma especificação mais detalhada das possíveis relações de causa e efeito entre os elementos da teoria. No exemplo a seguir, um investidor de impacto adquiriu um grupo de escolas privadas para estudantes de baixa renda e estabeleceu um conjunto de atividades pedagógicas e de gestão para aumentar o aprendizado e a retenção dos alunos. Observe, em particular, a forma de escrita dos componentes da teoria de mudança. As atividades identificam ações que os gestores se propõem a realizar (note os verbos "implantar", "adotar" e outros). Os produtos envolvem, em geral, mudanças em números e quantidades. Os resultados, por sua vez, identificam como a população-alvo se desenvolve em função das intervenções ("alunos *aprendem* os tópicos curriculares").

Na definição da teoria da mudança, é importante levar em consideração possíveis transformações positivas e negativas. Desconsiderar possíveis resultados negativos traz o risco de possíveis resultados adversos não intencionais gerados pela intervenção. Por exemplo, um projeto de geração de renda em uma comunidade rural pode estimular novas atividades agrícolas que permitam aumentar as receitas dos pequenos produtores, porém, ao mesmo tempo, gerar possíveis resultados negativos ao meio ambiente. Certas atividades podem também envolver risco de violação de direitos humanos básicos, como, por exemplo, discriminação de gênero ou trabalho em condições precárias. Uma sugestão é considerar inter-relações entre os objetivos mais gerais relacionados ao projeto (como discutido na Etapa 1), a fim de ter uma visão mais ampla de todos os possíveis resultados que podem ser causados pela intervenção.

A teoria de mudança permite também **definir um painel de indicadores de acompanhamento dos beneficiários do projeto. Porém, sob uma ótica de adicionalidade, a teoria de mudança, em si, não avalia o impacto do projeto.**

BOX 2. Exemplo de teoria de mudança para uma rede de ensino médio privada voltada a alunos de baixa renda.



5. Definição das métricas

Considerando-se os resultados das etapas anteriores, devem ser propostos indicadores para o projeto. Ainda que indicadores possam ser coletados para todas as etapas da teoria da mudança, inclusive para fins de monitoramento das ações e produtos, a análise de impacto coloca ênfase em indicadores de **resultado**. Como forma de preservar foco e facilidade de compreensão da análise, recomenda-se escolher poucas métricas de resultado que apresentem elevada relevância para avaliar o impacto gerado. Esses indicadores devem ser intimamente ligados às melhorias potenciais identificadas na teoria de mudança (Etapa 4).

Métricas ideais apresentam as seguintes características⁷:

- Têm elevada **relevância** em termos de geração de impacto isto é, representam resultados altamente desejáveis para a população-alvo. Se a métrica deriva da teoria da mudança realizada na Etapa 4, e se a teoria corretamente identifica resultados desejáveis na população, então ela deve ser relevante.
- São **passíveis de ser afetadas** pela ação dos executores do projeto. Não podem, por exemplo, ser indicadores muito amplos, cujo resultado dependa de muitos fatores sem

7. Essas características são particularmente relevantes tendo em vista que a avaliação de impacto pode também ser usada para definir metas para a gestão; consulte ROBERTS, J. Designing incentives in organizations. *Journal of Institutional Economics*, v. 6, n. 1, p. 125-132, 2010. Características de boas métricas são também discutidas em KUSEK, J.; RIST, R. C. *Ten steps to a results-based monitoring and evaluation system*. World Bank, 2004. Disponível em: <https://openknowledge.worldbank.org/bitstream/handle/10986/14926/296720PA-PER0100steps.pdf?sequence=1>. Acesso em: 02 dez. 2019.

controle direto dos executores. Em geral, métricas ligadas a resultados das atividades do projeto são mais passíveis de ação do que métricas mais ligadas a transformações sociais amplas.

- Devem ser medidas e verificadas com relativa **precisão**. Em geral, indicadores baseados em dados objetivos são preferíveis a indicadores muito subjetivos e com elevado erro de medição. Da mesma forma, deve-se cuidar para que os dados não sejam de fácil manipulação por gestores, apoiadores e outras partes interessadas em reportar resultados positivos. Nesse sentido, dados coletados por terceiras partes (como institutos independentes) e amplamente usados por atores no setor do projeto tendem a ser preferíveis.

- Devem ter **baixo custo** de medição — considerando-se, inclusive, o custo de obter dados em populações utilizadas como comparação (ou seja, indivíduos do grupo de controle). O custo da medição tende a se reduzir com o uso de dados secundários, publicamente disponíveis, em comparação com dados primários, coletados especificamente para o projeto.

Como sugestão, ao se considerar diferentes medidas alternativas, pode-se avaliar cada métrica com base nos itens acima e escolher aquelas que melhor atendam à maior parte desses critérios. Infelizmente, muitas vezes não é possível atender a todos os critérios, devendo os gestores optar por métricas que melhor se encaixem nos objetivos do projeto, e também métricas consolidadas e validadas na atividade de interesse. A teoria da mudança ajudará a identificar os resultados mais importantes a serem acompanhados, e caberá ao avaliador buscar os instrumentos de coleta mais adequados a seu caso.

Considerando as características de boas métricas acima discutidas, o **BOX 3** apresenta a ferramenta de **menu de métricas**, permitindo comparar métricas alternativas,

mostrando visualmente suas vantagens e desvantagens. Recomenda-se prosseguir para as etapas subsequentes de análise de impacto, após a escolha de métricas cujos atributos sejam, na sua maior parte, favoráveis, ainda que todos os indicadores de resultado possam ser acompanhados para fins de monitoramento do projeto.

O exemplo do **BOX 3** ilustra a escolha de possíveis métricas para um parque em uma área de conservação florestal e com atrações turísticas. O governo está interessado em estabelecer metas contratuais de impacto para guiar as ações de uma empresa privada escolhida para gerir o parque. Há, em particular, uma preocupação em gerar desenvolvimento para as comunidades nos municípios do entorno, além de indicadores de preservação ambiental. Na análise a seguir, as métricas ligadas à renda são relevantes, precisas e de baixo custo, porém a métrica de ganho de renda ligada a atividades de turismo é a mais passível de ação, pois a empresa gestora pode contratar pessoas da região para trabalhar no parque ou estimular empreendedores a ofertar serviços complementares (como hotéis e restaurantes). O indicador de satisfação dos usuários, apesar de relevante e passível de ação, tem um maior custo e é menos preciso — além de se basear em percepções subjetivas, traz risco de manipulação (por exemplo, gestores podem influenciar que a coleta seja feita em momento de menor lotação do parque, reduzindo-se a incidência de queixas). O indicador referente ao percentual de área seguindo regramentos ambientais, se informado pela empresa gestora, pode exigir contratação de auditoria técnica especializada, o que tende a aumentar os custos. A partir dessa análise, conclui-se que o segundo indicador de resultado (renda em atividades ligadas ao turismo) é um candidato potencial para seguir nas próximas etapas de avaliação de impacto por meio de cálculo de adicionalidade, ainda que outros indicadores também possam ser acompanhados para fins de monitoramento do projeto.

Recomenda-se escolher poucas métricas de resultado que apresentem elevada relevância para avaliar o impacto gerado.

Esses indicadores devem ser intimamente ligados às melhorias potenciais identificadas na teoria de mudança.

BOX 3. Menu de métricas: exemplo de análise de métricas alternativas para um projeto de parque turístico

RESULTADO A SER MEDIDO	MÉTRICA	RELEVANTE?	PASSÍVEL DE AÇÃO?	PRECISA?	DE BAIXO CUSTO?
1. Comunidades no entorno do parque obtêm maior renda e bem-estar	Aumento anual de renda <i>per capita</i> dos municípios do parque e no seu entorno.	✓ Desenvolvimento local é um dos objetivos do governo com o projeto.	✗ A economia local pode ser afetada por outros fatores, além de pelas atividades do parque.	✓ Há estatísticas disponíveis e confiáveis sobre renda local.	✓ As estatísticas são amplamente divulgadas pelo governo e de fácil acesso.
2. Comunidade local aumenta a sua renda com atividades ligadas ao turismo	Aumento anual de renda <i>per capita</i> em serviços de hotelaria, restaurantes e outros nos municípios do parque.	✓ Desenvolvimento local é um dos objetivos do governo com o projeto.	✓ O parque deve influenciar a movimentação de hotéis e restaurantes, por exemplo.	✓ Há estatísticas disponíveis e confiáveis sobre renda local.	✓ As estatísticas são amplamente divulgadas, de fácil acesso e desagregadas por setor de atividade.
3. Usuários do parque ficam mais satisfeitos	Nota média de satisfação dos usuários, a partir de pesquisas conduzidas por institutos contratados.	✓ Uma maior satisfação pode indicar a qualidade dos serviços aos visitantes.	✓ Esforço para prover melhores serviços tende a afetar positivamente a satisfação.	✗ Indicadores de satisfação são mais subjetivos, e há risco de manipulação.	✗ Elevado custo para realização de pesquisas e dificuldade de obter dados de comparação.
4. Parque atende a regramentos de conservação	Percentual das áreas do parque seguindo regramentos de conservação.	✓ Estando em área de conservação, ações de preservação ambiental são centrais.	✓ Gestores do parque podem controlar diretamente o estado das áreas de conservação.	✓ Há protocolos claros indicando aderência a padrões de conservação.	! Gestores realizam esse monitoramento, mas pode ser necessária validação técnica externa.



6. Definição do método de verificação de adicionalidade e do nível de medição

Após a definição das métricas, é importante definir como será feita a verificação da adicionalidade, isto é, considerar o que provavelmente teria acontecido com a população-alvo, caso o projeto não tivesse sido executado. Nessa etapa, são definidos o grupo dos **tratados** (indivíduos e comunidades beneficiadas pelo projeto) e o grupo de **controle** (indivíduos e comunidades que não receberam as intervenções). O grupo de controle deve ter características similares às

do grupo tratado, de forma a garantir que ele represente adequadamente o que poderia ter acontecido aos tratados sem o projeto.

A forma de verificação de adicionalidade varia de acordo com o **nível de medição** definido para o projeto, conforme detalhado na seção seguinte. Níveis mais elevados de medição trazem mais confiança de que mudanças na população-alvo foram realmente causadas pelo projeto, porém tendem a ser

mais complexos e custosos. Nessa etapa, especialmente caso sejam adotados níveis de medição mais elevados, recomenda-se o apoio de especialistas para ajudar na identificação das técnicas mais apropriadas e suas potenciais limitações.

Uma vez definida a forma de verificação de adicionalidade, e com base nos indicadores de resultado escolhidos, é importante definir com precisão como será calculado o impacto. No exemplo da **Figura 1**, o impacto será medido como a diferença de variação de renda entre a comunidade tratada e a de controle, considerando-se os dois períodos definidos. Essa medição permite aferir o

tamanho do impacto total, definido como o ganho médio por indivíduo vezes o número de indivíduos presentes no grupo tratado.

A medição também evolui uma definição do **prazo** para se aferir o impacto. Muitos projetos de impacto têm um prazo razoavelmente longo, em função das características de atividade e da necessidade de se implementar um conjunto de ações complexas e complementares. A Etapa 3, de *benchmarking*, permite ainda revisar estudos prévios, avaliando-se não apenas o quanto determinadas intervenções podem gerar de impacto, mas também em quanto tempo.



7. Plano Amostral

A definição do tamanho amostral — quantas pessoas ou quantos grupos participarão da avaliação do projeto — é outra crucial etapa do plano de medição. Essa etapa é particularmente importante no caso de níveis de medição mais elevados, empregando técnicas que se propõem a identificar de forma mais precisa o efeito do projeto sobre a população-alvo. Com um adequado tamanho amostral, é possível verificar com maior precisão estatística a diferença entre o resultado medido antes e depois do projeto, bem como a diferença entre os grupos tratados e de controle.

Essencialmente, um maior tamanho da amostra reduz o risco de as análises não encontrarem um efeito positivo do projeto, supondo que realmente ele exista. Em geral, o tamanho da amostra depende do tamanho do efeito esperado: tudo o mais constante, quanto menor o impacto esperado do projeto, maior o tamanho necessário da amostra para que se possa medi-lo. No entanto, os custos de realizar o tratamento e de coletar os dados relevantes devem ser levados em consideração para otimizar a amostra⁸.

8. Uma discussão técnica sobre como definir o tamanho da amostra pode ser vista no texto de DUFLO, E.; GLENNISTER, R.; KREMER, M. Using randomization in development economics research: A toolkit. In T. P. Schultz, & J. Strauss (Eds.), **Handbook of Development Economics**: p. 3895–3957. Amsterdam: Elsevier, 2008. Uma ferramenta prática para cálculo de tamanho amostral pode ser obtida em: Optimal Design Software. Disponível em: <https://sites.google.com/site/optimaldesignsoftware/home>. Acesso em: 6 mai. 2019.



8. Cronograma de medição

Nesta etapa, é importante definir o período de coleta de dados, indicando, se for o caso, as medições antes e depois do início das intervenções (como indicado na **Figura 1**). Podem também ser estabelecidas medições intermediárias, como forma de acompanhar a evolução do projeto. Como discutido anteriormente, os gestores do projeto devem ter consciência de que, na maioria dos casos, a verificação de impacto necessita de um tempo razoável, depois da intervenção. Por exemplo, no setor de educação, em que o aprendi-

zado pelos alunos requer um esforço longo e persistente, o horizonte de tempo tende a ser longo, principalmente se houver interesse em medir a conclusão de um ciclo educacional ou seu impacto sobre renda futura.

A decisão sobre o período ideal para acompanhar e medir os resultados do projeto pode ser informada por estudos prévios avaliando o impacto de intervenções similares ao longo do tempo (como discutido na Etapa 3, de *benchmarking*). ■

NÍVEIS DE MEDIÇÃO

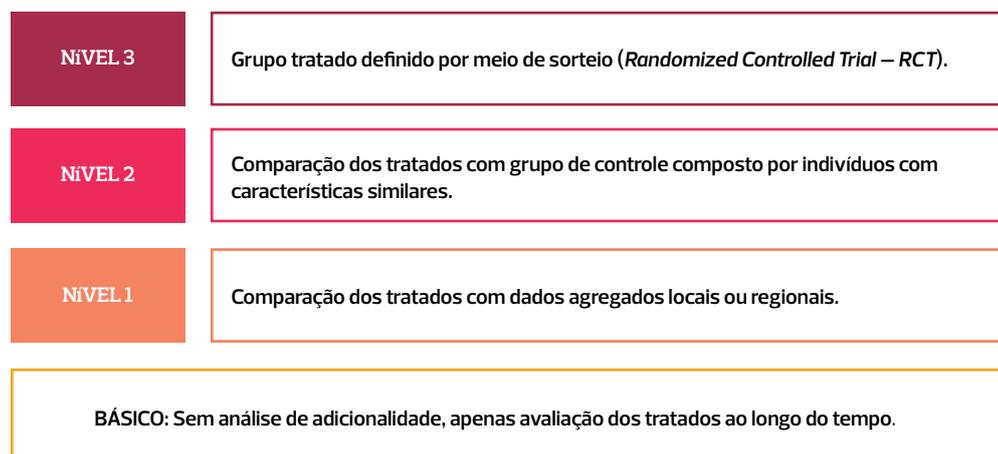
Os níveis de medição, essencialmente, propõem formas alternativas de corrigir o efeito de fatores externos decorrentes de diferenças preexistentes entre os grupos tratados e de controle. Por exemplo, imagine que um investidor social queira avaliar o efeito de um novo método pedagógico e ofereça esse método a diversas escolas. É possível que as escolas que aceitem aderir ao programa sejam justamente aquelas com gestão mais motivada para melhorias no ensino. Se essa variável afetando a escolha ("gestão motivada") não for observada pelo avaliador, pode-se concluir erroneamente que possíveis melhorias nas escolas que voluntariamente aderiram ao programa resultaram do programa em si, ainda que a causa real tenha sido a propensão natural das escolas escolhidas a melhorar o seu ensino.

Para minimizar esse erro, especialistas em avaliação com verificação de adicionalidade enfatizam que, idealmente, os grupos e indivíduos tratados devem ser escolhidos de forma aleatória, por sorteio. No entanto, sabe-se que o sorteio de indivíduos ou grupos é particularmente complexo e, em muitos casos, até mesmo inexecutável. A decisão pelo investimento em determinados projetos, seja esta tomada por fundos ou indivíduos, seja por empresas ou governos, na maior parte dos casos se-

gue pressupostos e planos previamente estabelecidos. Assim, em muitos casos, os apoiadores e executores de projetos de impacto acabam tendo de recorrer a outros métodos que tentam corrigir diferenças prévias entre os grupos tratado e de controle.

Qualquer que seja a forma de verificação escolhida, é importante manter total **transparência** sobre o método adotado e suas possíveis limitações. Especificamente, abaixo são propostos **níveis (tiers) de medição**, de acordo com o grau de robustez pretendido para a avaliação de impacto, em especial no que tange ao rigor sobre a aferição do efeito *causal* do projeto. Na **Figura 3**, partindo-se da base rumo ao topo, inicia-se com um tipo de medição denominado de *básico* – que não envolve análise de adicionalidade – como forma de expor como cada nível permite maior rigor na estimativa do efeito do projeto. Uma vez que os gestores normalmente utilizam mais de uma métrica, é possível que os próprios indicadores alternativos sejam avaliados de acordo com níveis de medição distintos. É provável que a maior parte dos indicadores escolhidos seja acompanhada de forma básica, enquanto alguns outros poucos indicadores possam ser medidos utilizando-se níveis de medição mais elevados.

Figura 3: Níveis de medição para verificação de adicionalidade



BÁSICO: Mensuração apenas dos resultados do grupo tratado

Essa abordagem *não* envolve cômputo de adicionalidade e *não* deve ser considerada como um nível de medição, uma vez que não inclui análise de cenário contrafactual (isto é, do que teria acontecido com a comunidade-alvo sem a intervenção). Entretanto, para muitos casos, é a forma mais factível de ser executada, uma vez que é mais fácil coletar dados dos indivíduos ou grupos participando do projeto. Nesse caso, simplesmente observa-se como determinados indicadores variaram no início e ao longo das intervenções. É tam-

bém uma abordagem bastante usada para acompanhar um painel de indicadores de desempenho do próprio projeto, como discutido na Etapa 4. Contudo, ao observar variações positivas nesses indicadores, os gestores devem ter o cuidado de não necessariamente inferir que essas mudanças foram causadas pelas intervenções realizadas. A própria população-alvo pode já ter uma tendência de melhoria, mesmo antes do início das intervenções, e podem ter ocorrido mudanças externas favorecendo os resultados do projeto.

Nível 1: Mensuração comparando o grupo tratado com dados agregados locais ou regionais

Nesse nível, empregam-se dados agregados como forma de comparação dos resultados do projeto ao que poderia ter ocorrido sem a sua implantação. Antes da medição, há uma definição deliberada da localidade onde o projeto será realizado e de quem são os indivíduos beneficiados. Além disso, para fins comparativos, são utilizados dados agregados já disponíveis para uma determinada região onde seja feito o investimento. Devem ser buscados dados que permitam avaliar o que ocorreu com o projeto compa-

rativamente ao resultado agregado, antes e depois da intervenção (BOX 4).

Nesse nível, deve-se, ainda, levar em conta o tamanho da intervenção. Se cobrir um grande número de indivíduos em determinada região, os dados agregados estarão sensivelmente afetados por ela, e a comparação ficará prejudicada. Uma possibilidade nesse caso é buscar a comparação com outros locais próximos, de características semelhantes.

BOX 4. Exemplo de verificação de adicionalidade no Nível 1

Uma empresa agroindustrial decide realizar compras de produtos agrícolas de pequenos produtores familiares em certos municípios de baixa renda, ao mesmo tempo apoiando esses produtores com assistência técnica e gerencial. A definição das comunidades beneficiadas é feita de forma deliberada pelo executor do projeto. Utilizando a renda dos produtores como uma métrica de impacto, a avaliação é efetuada comparando-se a renda das famílias apoiadas frente à renda de famílias em áreas rurais do município ou da microrregião onde foi feito o investimento, antes e depois da intervenção. Será verificada adicionalidade no Nível 1 caso a variação de renda das famílias apoiadas supere a variação agregada das famílias na região, no mesmo período.

Nível 2: Mensuração utilizando um grupo de controle com características semelhantes às do grupo tratado

Nesse nível, embora ainda não seja implementado o sorteio, há a construção de um grupo de controle formado por indivíduos ou comunidades comparáveis aos tratados. Por isso, as técnicas de verificação de adicionalidade no Nível 2 são normal-

mente chamadas de **não experimentais**. Da mesma forma que no Nível 1, há uma definição de quem são os indivíduos beneficiados. E, no processo de avaliação, empregam-se técnicas para construir um grupo de controle formado por indivíduos

que sejam parecidos com os indivíduos tratados. Ou seja: em vez de utilizar dados agregados de um determinado local, há uma tentativa de acompanhar grupos ou indivíduos que não recebem o investimento, porém têm características similares às dos tratados. Uma forma bastante comum de medir impacto no Nível 2 é utilizar o método das **diferenças-em-diferenças**, exemplificado na **Figura 1**. Nesse método, o impacto é medido como a evolução dos resultados dos tratados em comparação à evolução medida no grupo de controle. A evolução do grupo de controle essencialmente serve como uma estimativa do cenário contrafactual: o que teria acontecido à população-alvo do projeto sem a intervenção que foi realizada.

Vale lembrar, contudo, que o Nível 2 requer que os grupos de tratamento e controle sejam realmente comparáveis, tanto em termos de seus atributos quanto em termos da própria evolução passada do seu desempenho. Assim, no método das diferenças-em-diferenças, um importante aspecto é que os grupos tratado e de controle exibam uma **tendência** similar de evolução da métrica de impacto antes da intervenção. Na **Figura 1**, a comunidade tratada e a de controle evoluíram “paralelamente” antes do projeto se iniciar, o que sugere que elas eram afetadas por fatores similares. Tendências não paralelas poderiam sugerir a existência de fatores distintos influenciando a evolução dos grupos.

Outra forma muito comum de escolher indivíduos com características semelhantes entre os grupos de tratados e de controle

envolve a chamada técnica de **pareamento**. Nela, com base em características observáveis relevantes para a seleção de participantes da intervenção (como idade, sexo, renda e assim por diante), procura(m)-se uma ou mais unidade(s) no grupo de controle que sejam as mais parecidas quanto possível com cada unidade no grupo de tratamento (veja o exemplo 5.1 do **BOX 5**). A técnica de pareamento é também usualmente combinada com a de diferenças-em-diferenças.

Em casos em que exista competição para participar do projeto e um determinado critério de seleção, uma possibilidade é considerar a **descontinuidade** gerada pela própria seleção. Isto é, podem ser comparados os participantes que quase não foram aceitos (tratados) àqueles que quase foram aceitos, mas que, eventualmente, ficaram abaixo do limite de corte (controle). Por estarem muito próximos do limite, é bastante provável que esses participantes sejam muito similares em suas várias características (veja o exemplo 5.2 do **BOX 5**).

Ainda em outras situações em que se observa o impacto em um caso específico (uma única organização ou município, por exemplo), há a possibilidade de criar um **controle sintético**, a partir de uma combinação de outros casos que mais se aproximem do caso tratado. Nessa técnica, é preciso ter dados do caso tratado e de potenciais outros casos antes do início do projeto, de forma a criar um controle sintético que adequadamente consiga replicar a evolução do caso que recebeu a intervenção⁹.

9. Um exemplo de forma de implementação de técnica de pareamento pode ser visto em: ABADIE, A.; DRUKKER, D.; HERR, J. L.; IMBENS, G. W. Implementing matching estimators for average treatment effects. *The Stata Journal*, v. 4, n. 3, p. 290–311, 2004. Para uma discussão mais geral de vários métodos alternativos, consulte: ANGRIST, J.; PISCHKE, J.-S., **Mostly Harmless Econometrics**, Princeton: Princeton University Press, 2009 e MENEZES FILHO, N. A. (editor). **Avaliação Econômica de Projetos Sociais**. São Paulo: Dinâmica Gráfica e Editora, 2012. A técnica de controle sintético é discutida em ABADIE, A.; DIAMOND, A.; HANMUELLER, J. Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco control program. *Journal of the American Statistical Association*, v. 105, n. 490, p. 493–505, 2010.

BOX 5. Exemplos de verificação de adicionalidade no Nível 2.

Exemplo 5.1: Em um projeto de impacto cujo indicador é a renda de famílias, a avaliação é efetuada comparando-se a renda das famílias tratadas com dados sobre a renda de famílias no grupo de controle. A definição das comunidades beneficiadas pelo projeto é feita de forma deliberada pelo executor. Neste nível de medição, as famílias de controle são então selecionadas com base em características observáveis mais semelhantes quanto possível àquelas das tratadas. Por exemplo, no grupo de controle podem ser selecionadas famílias com níveis iniciais de renda, escolaridade e número de filhos similares aos níveis observados nas famílias tratadas.

Exemplo 5.2: Uma organização pretende implementar um programa de microcrédito, convidando microempreendedores a participar do programa, pré-avaliados com base em indicadores de risco de crédito (renda, histórico em empréstimos anteriores etc.). Havendo vagas limitadas, apenas os mais bem-avaliados serão selecionados, a partir de um critério de corte estabelecendo um nível mínimo de risco de crédito. No limiar desse critério de corte, pode-se supor que os indivíduos são parecidos, dado que apresentam avaliação similar de risco de crédito. Dessa forma, pode-se comparar o grupo de microempreendedores selecionados logo acima da linha de corte com aqueles que não foram selecionados, mas que ficaram logo abaixo da linha.

Nível 3: Mensuração com sorteio (*Randomized Controlled Trial – RCT*)

As técnicas do Nível 3, também chamadas de **experimentais**, permitem máxima confiabilidade da estimativa de avaliação de impacto. Neste nível, a decisão de quais indivíduos ou comunidades serão sujeitos à intervenção é feita de forma totalmente aleatória, logo, reduzindo possíveis vieses de seleção dos grupos-alvo do projeto. Em casos em que não seja possível ou desejável excluir determinados grupos, pode-se implementar um desenho alternativo em que um grupo seja encorajado a adotar o que é proposto pela intervenção, sendo a atribuição do encorajamento definida de forma aleatória entre os indivíduos ou as comunidades-alvo (**BOX 6**).

A escolha por sorteio, se bem-feita, garante que não existam diferenças entre as características, observáveis ou não, dos grupos de tratados e de controle. Por isso, medições do Nível 3 em geral não necessitam de medição em um período inicial, anterior ao projeto, como indicado na **Figura 1**. Seguindo-se procedimentos adequados de sor-

teio, pode-se avaliar o desempenho comparativo dos grupos tratados e de controle em um momento após o início do projeto.

No Nível 3, é particularmente importante atentar nos cálculos de tamanho mínimo da amostra, conforme discutido na Etapa 7 da seção anterior. É importante, contudo, não confundir amostragem aleatória com definição aleatória de tratamento (por sorteio). No Nível 3, o sorteio se dá na escolha do grupo que deve receber ou não a intervenção. Por exemplo, suponha que determinado projeto de educação tenha beneficiado 30 escolas, sendo o processo de escolha destas definido de forma não aleatória. Então, o Nível 3 não será atingido mesmo se forem selecionadas, aleatoriamente, outras 30 escolas como grupo de controle. Isso porque a seleção de quem recebeu a intervenção não foi feita de forma aleatória, podendo, portanto, ser afetada por fatores não observáveis pelos avaliadores no momento de calcular o impacto da intervenção. ■

10. Nesse caso, pode-se medir o impacto da intervenção a partir do estimador de efeito médio local do tratamento (LATE – *Local Average Treatment Effect*). Ver: ANGRIST, J.; PISCHKE, J.-S., op. cit.

BOX 6. Exemplos de verificação de adicionalidade no Nível 3.

Exemplo 6.1: Uma organização oferece um programa limitado de microcrédito em um contexto em que não há recursos para apoiar todos da comunidade. A organização pode convidar interessados e definir quem vai receber o microcrédito, de forma aleatória. Alternativamente, pode emprestar para aqueles com melhor risco de crédito e, então, aleatorizar dentro do conjunto daqueles que quase foram selecionados, isto é, logo abaixo da linha de corte de risco de crédito. Considerando-se esse conjunto, pode-se definir de forma aleatória um grupo de indivíduos tratados que receberão o crédito, sendo então o grupo de controle definido por aqueles não escolhidos por esse procedimento de sorteio.

Exemplo 6.2: Um amplo grupo de alunos de uma rede pública recebe acesso a uma plataforma de aprendizado por internet. Ainda que todos os alunos tenham acesso à plataforma, pode-se sortear um grupo de estudantes que receberá mensagens estimulando o seu uso. Considerando-se que uma fração desses alunos que recebeu as mensagens usará de fato a ferramenta (ou seja, somente estes serão efetivamente tratados), é possível comparar o desempenho deles ao desempenho do grupo de controle de alunos que não recebeu as mensagens. Essa análise, entretanto, requer ajustes estatísticos, uma vez que se quer avaliar quem efetivamente usou a plataforma, não necessariamente quem recebeu ou não recebeu o estímulo¹⁰.

COMPARANDO OS NÍVEIS DE MEDIÇÃO

O **Quadro 1** apresenta um comparativo das vantagens e limitações de cada nível de medição discutido anteriormente. Em linhas gerais, movendo-se em direção ao Nível 3, é possível aumentar a confiabilidade das medições de impacto, no sentido de garantir que

o projeto de fato causou as mudanças detectadas na população-alvo. Porém, ao mesmo tempo, aumenta-se a complexidade das análises e do desenho da pesquisa, além de se exigir potencialmente mais recursos técnicos e financeiros para as análises.

Quadro 1 – Vantagens e limitações dos níveis de medição

	VANTAGENS	LIMITAÇÕES
Básico	<ul style="list-style-type: none">▪ Dados mais facilmente disponíveis (coletados do próprio projeto e da sua população-alvo).▪ Maior simplicidade nas análises (comparação apenas dentro do grupo tratado, antes e depois).	<ul style="list-style-type: none">▪ Não envolve verificação de adicionalidade, pois não há análise de cenário contrafactual (isto é, do que teria acontecido com a população-alvo sem o projeto).
Nível 1	<ul style="list-style-type: none">▪ Em muitos casos, gestores têm à disposição apenas dados agregados (por exemplo, renda municipal, em vez da renda das famílias do município).▪ Comparada a outros métodos de verificação de adicionalidade, a análise é mais simples (resultados do projeto versus resultados agregados).	<ul style="list-style-type: none">▪ As características da população agregada podem ser muito distintas daquelas da população-alvo, ainda que dentro do mesmo recorte geográfico.▪ Não permite estimação de efeito causal, dado que não envolve técnicas estatísticas para controle de potenciais diferenças entre grupos.
Nível 2	<ul style="list-style-type: none">▪ Em muitos casos, a população-alvo tratada já foi definida, sendo inviável realizar o sorteio.▪ Ao se construir grupos de controle com características similares às dos tratados, evita-se viés gerado por diferenças substanciais entre os grupos.▪ As diferentes técnicas do Nível 2 permitem realizar inferência estatística sobre o impacto observado no grupo tratado, frente ao grupo de controle.	<ul style="list-style-type: none">▪ As técnicas do Nível 2 só permitem avaliar o efeito causal do projeto se não houver efeito relevante de outros fatores não observados nem medidos, o que é difícil de verificar na prática.▪ É preciso obter informações de características não apenas daqueles afetados pelo projeto, mas também daqueles no grupo de controle.▪ Deve haver casos no grupo de controle que sejam próximos, em suas características, aos tratados. Caso contrário, não é possível realizar a comparação.
Nível 3	<ul style="list-style-type: none">▪ Ao se realizar o sorteio, é possível garantir com mais segurança que o efeito medido foi realmente causado pelo projeto.▪ Em geral, o sorteio não requer medição antes do início do projeto; os indivíduos e grupos podem ser comparados após a realização das intervenções.	<ul style="list-style-type: none">▪ A técnica exige atenção particular para o tamanho da amostra. Torna-se inviável em contextos com poucos casos recebendo a intervenção.▪ Desenhos experimentais no Nível 3 são muito sujeitos a problemas de efeitos cruzados, atrito e outros, discutidos na seção seguinte do <i>Guia</i>.▪ Podem surgir dilemas éticos em casos em que não se queira excluir indivíduos de uma determinada intervenção.

CUIDADOS DURANTE A MEDIÇÃO

Algumas precauções são extremamente importantes para a realização satisfatória de uma medição de impacto. Abaixo, são descritos de forma pontual alguns aspectos críticos que o executor do projeto deve considerar na realização do projeto de medição¹¹.

EXTERNALIDADES

Um dos principais pontos a ser considerado é que deve existir o mínimo possível de **externalidades** ou "efeitos cruzados" entre os indivíduos de grupos diferentes. Por exemplo, considere que se pretenda avaliar o impacto de um projeto empresarial para o aumento de renda em uma determinada localidade. Para tanto, previamente ao projeto, realiza-se um levantamento de dados nessa comunidade e na comunidade vizinha, que não foi considerada para o projeto. A externalidade pode ocorrer se, por exemplo, a comunidade tratada, ao receber renda advinda do investimento, puder despender parte dos seus recursos comprando produtos ou transferindo renda à outra comunidade. Se

isso ocorrer, a adicionalidade de impacto do programa seria subestimada.

Essas externalidades são problemáticas para avaliações de impacto, pois podem distorcer o resultado, pela "contaminação" do grupo de controle com a intervenção feita no grupo tratado. Para contornar esse problema, é importante garantir que não haja comunicação, troca de informação, nem troca física de bens e serviços entre os indivíduos. Uma das formas de conseguir isso é garantir que os grupos tratado e de controle estejam geograficamente distantes ou isolados, com menor risco de efeitos cruzados.

11. Esta seção é fortemente baseada em DUFLO, E.; GLENNERSTER, R.; KREMER, M., op. cit.

ATRITO

Dado o tempo entre a realização da intervenção e a avaliação, é possível que ocorra um "descasamento" entre os indivíduos observados no momento inicial e os avaliados no momento final, tanto no grupo de controle quanto no grupo tratado. Por exemplo: pode ocorrer que alguns indivíduos observados no momento inicial não residam mais na localidade, para avaliação final. Vale destacar que, apesar da importância de se realizar sorteio para a presente metodologia, ela não garante a inexistência de atrito não aleatório ao fim. Ademais, o problema do atrito reforça a importância de se definir, anteriormente à execução do projeto, uma amostra suficientemente grande para a realização das medições.

Como exemplo prático desse problema, considere um projeto de impacto para o aumento de renda em uma determinada localidade. Ainda que, inicialmente, os gru-

pos tenham sido escolhidos de forma aleatória, ao fim do projeto pode-se deparar com atrito não aleatório, decorrente da não localização de indivíduos que não tenham recebido o aumento de renda, frente à localização de indivíduos que o tenham recebido e continuem na localidade. Nesse caso, o atrito gera um viés na medição de impacto, pois haverá uma perda não aleatória de dados dos indivíduos do grupo de controle.

É sempre importante expor as taxas de atrito presentes tanto no grupo tratado quanto no de controle, de forma transparente. Ou seja, ao final da medição, é preciso evidenciar quantos indivíduos medidos no primeiro momento não foram encontrados para a medição no segundo momento, assim como comparar as características médias dos grupos antes e depois do atrito, tanto no grupo de controle quanto no grupo tratado.

Externalidades podem distorcer a medição de impacto pela "contaminação" do grupo de controle com a intervenção feita no grupo tratado.

EFEITO HAWTHORNE¹² E EFEITO JOHN HENRY¹³

Por fim, outra das principais limitações que podem ocorrer em uma medição de impacto é a mudança de comportamento dos indivíduos tanto do grupo tratado, quanto do grupo de controle. O chamado **efeito Hawthorne** pode ocorrer pela percepção dos indivíduos tratados de que estão sob algum tipo de intervenção e, assim, há uma mudança de comportamento em resposta. Em projetos de educação, por exemplo, pode haver um esforço superior dos professores e dos alunos, pela conscientização desses indivíduos de que seu grupo está sendo beneficiado pela intervenção.

Já o efeito comportamental no grupo de controle, chamado de **efeito John Henry**, pode ocorrer quando professores do grupo de controle se sentem desafiados e começam a competir com os do grupo tratado, para mostrar que também teriam condições de se beneficiar do programa. O problema pode ocorrer também em outra direção: por não receberem tratamento, os professores se desmotivam, gerando uma superestimação do real impacto do projeto. Nos dois casos, o efeito do programa pode ser distorcido.

Ainda que respostas comportamentais sejam sempre complexas, há formas de desenho de medição que possibilitam minimizar os efeitos Hawthorne e John Henry. Por exemplo, podem-se utilizar três grupos: tratado, de controle e **placebo**. Neste último, se estabelece um grupo que será observado ou que manterá relações com o executor do projeto social, sem, contudo, ser efetivamente beneficiado pelo projeto.

Considere, por exemplo, um projeto no qual uma grande empresa deseja melhorar a renda de determinadas comunidades por meio da compra de insumos locais. Assim, a empresa não deve apenas considerar uma comunidade de atuais fornecedores (tratada) e uma comunidade comparável de não fornecedores (controle). A empresa pode também monitorar uma comunidade placebo de não fornecedores com a qual tenha alguma forma de relação (por exemplo, a empresa pode enviar gestores para conversar com essa comunidade, mas sem estabelecer contratos de suprimento). O impacto medido no grupo tratado deve ser superior ao verificado nos grupos de controle e placebo. ■

12. Efeito Hawthorne leva esse nome em função da experiência homônima ocorrida em 1927 na fábrica West Electric Company (Chicago, EUA), onde se verificou que os indivíduos mudam seu comportamento de acordo com alterações percebidas no ambiente de trabalho.

13. O efeito John Henry é baseado na história de um trabalhador em um local onde se introduziu, ainda em fase de testes, uma nova ferramenta que visava a aumentar produtividade. Na ocasião, o trabalhador se sentiu desafiado a trabalhar mais, de forma a mostrar que poderia superar a nova tecnologia.

AVALIANDO O RESULTADO ECONÔMICO DO PROJETO

O presente *Guia* coloca foco em procedimentos para se estimar o impacto de um determinado projeto, tendo em vista objetivos inicialmente definidos e resultados elencados pela teoria de mudança. Assim, o *Guia* não se propõe a discutir como computar o retorno econômico da intervenção socioambiental em função de quanto se gastou para realizá-la em comparação aos benefícios gerados aos participantes do programa.

Não há forma única para se calcular o retorno econômico do projeto social, principalmente quando o benefício for medido em termos de indicadores não monetários. Porém, é importante perguntar se, com o montante investido, seria possível obter resultados mais expressivos — ou, visto de outra forma, se o mesmo resultado poderia ser obtido com menos recursos.

Uma possibilidade, nessa linha, é construir cenários plausíveis de como os indicadores de resultado reverterão em ganhos monetários, calculando-se assim a relação entre custo e impacto em termos monetários do projeto — ou seja, a sua **relação benefício-custo**. Em alguns casos, é possível comparar os indicadores sociais ou ambientais de dois projetos diferentes, mas cujos objetivos e montante de recursos empregados sejam os mesmos.

Por exemplo, um estudo mostrou que aumento de jornada escolar de quatro para cinco horas ou redução de tamanhos de classe de 38 para 30 alunos gera impactos similares nas notas dos estudantes. Porém, quando considerado o custo envolvido, o aumento de jornada apresenta a maior razão benefício-custo¹⁴.

Em casos onde os resultados dos projetos difiram, é preciso atribuir valores monetários a um determinado montante de impacto estimado. Por exemplo, um projeto educacional no Ensino Médio pode aumentar o aprendizado dos alunos e tornar mais provável o seu ingresso em uma boa faculdade, com reflexos positivos em ganho de renda futura. Esse ganho pode ser comparado ao custo do projeto.

Outra possibilidade é a chamada metodologia de disposição a pagar. Pergunta-se diretamente ao público de interesse quanto ele estaria disposto a pagar por determinado bem ou serviço. É o caso, por exemplo, da valoração de projetos culturais ou iniciativas que visem a aumentar a satisfação do público com uma determinada atividade (como visitas a parques ou museus). O “preço” reportado pelos usuários pode então ser comparado ao custo das atividades envolvidas no serviço. ■

14. Oliveira, J. M. Custo-efetividade de políticas de redução do tamanho da classe e de ampliação da jornada escolar: uma aplicação de estimadores de *matching*. **Relatório de pesquisa**, 11. Prêmio BNDES de Economia, 2010.

Dados Internacionais de Catalogação na Publicação (CIP)

I59g Insper Metricis

Guia de avaliação de impacto socioambiental para utilização em projetos e investimentos de impacto: guia geral com foco em verificação de adicionalidade. - 4a. ed. - São Paulo: Insper, 2020.

24 p. il.: col. 29,7cm

1. Investimentos de impacto 2. Avaliação de impacto
3. Verificação de adicionalidade (additionality)
4. Socioambiental 5. Gestão de projetos

I. Insper Metricis II. Título

CDU 504

Catálogo na Fonte: Ricardo Rodrigues Ramos CRB 8/9309

Insper METRICIS

Núcleo de Medição para Investimentos de Impacto Socioambiental