

Monitoramento e Avaliação de Políticas Públicas

Fábio Rodrigues de Moura

Avaliação de Impacto de Programas sociais

Introdução

- A decisão dos formuladores de políticas públicas de investir em programa social sempre parte do pressuposto de que a intervenção terá impacto positivo sobre um conjunto de resultados de interesse coletivo ou individual que, em última instância, deve trazer benefícios para, ao menos, uma parcela da sociedade.
- O principal propósito de uma avaliação de impacto é verificar se, na realidade, determinado programa está alcançando os objetivos, ou impactos, esperados.
- Entendemos por impacto as diferenças entre a situação dos participantes do projeto após terem participado e a situação em que estariam, caso não tivessem tido acesso a ele.

Introdução

- Dessa forma, o impacto do programa é definido como o contraste entre duas situações: uma real (a situação dos participantes após a participação no projeto) e outra hipotética (a situação em que estariam caso não tivessem tido a oportunidade de participar do programa).

Introdução

- Essa definição de impacto, entretanto, é apenas o ponto de partida.
- Avaliações de impacto certamente consomem recursos que, alternativamente, poderiam ser investidos diretamente no próprio programa.
- A crescente expansão de estudos dessa natureza sinaliza que a importância de dedicar esforços e recursos para esse fim é cada vez mais reconhecida.
- Mas, por que avaliações de impacto são importantes?

Introdução

- Os resultados dessas avaliações têm uma variedade de utilidades potenciais.
- A mais óbvia delas é utilizar as informações produzidas para a melhoria do desenho do próprio programa em questão.
- Além disso, os resultados produzidos podem ser usados para auxiliar no desenho de outros programas subsequentes na mesma comunidade ou até mesmo para informar formuladores de políticas que pretendam desenhar programas com objetivos similares em outras comunidades.

Por que avaliar o impacto de um programa social?

- Nossa sociedade investe grande quantidade de recursos, tanto humanos quanto financeiros, em programas sociais.
- Avaliações de impacto nos permitem verificar se esses recursos estão de fato sendo aplicados da melhor maneira possível.
- No entanto, avaliações de impacto também consomem recursos. Será que a realização da avaliação de impacto do programa é sempre o melhor uso de recursos ou seria melhor revertê-los em maior investimento no próprio programa? Afinal, os benefícios de uma avaliação de impacto superam seus custos?

Por que avaliar o impacto de um programa social?

a) Não é evidente que um programa amplamente utilizado tenha impacto?

- A justificativa da importância da aplicação de recursos para avaliar impacto não é tão óbvia. Uma contra-argumentação válida poderia ser fundamentada na alegação de que é pouco provável que um programa com acesso voluntário não tenha impacto sobre aqueles que o procuram e dele se utilizam.
- Por que alguém dedicaria tempo e esforço a um programa que não lhe traz benefícios e, portanto, não tem qualquer impacto?
- A ampla utilização de um programa por, ao menos, um segmento da sociedade não seria por si só indício da existência de impacto?

Por que avaliar o impacto de um programa social?

a) Não é evidente que um programa amplamente utilizado tenha impacto?

- De fato, é pouco provável que a maioria dos programas existentes não traga benefício algum a seus beneficiários. Afinal, em sua maioria, os programas sociais são voluntariamente utilizados pelos seus beneficiários que, portanto, devem estar convencidos da sua utilidade.
- Em sua maioria, os programas sociais se baseiam em teorias que preveem o elo entre o acesso ao programa e impactos positivos sobre um conjunto pré-determinado de resultados. Portanto, os formuladores de políticas públicas, geralmente, contam com argumentos teóricos e, frequentemente, com evidência empírica, ambos sinalizando a eficácia do programa em questão.

Por que avaliar o impacto de um programa social?

a) Não é evidente que um programa amplamente utilizado tenha impacto?

- Assim, investigar a existência de impacto de um programa cuidadosamente desenhado e voluntariamente utilizado pela sociedade não seria questionar a racionalidade dos usuários e a validade dos argumentos teóricos e empíricos adotados no desenho do programa?
- Existem duas possíveis respostas a essa pergunta.

Por que avaliar o impacto de um programa social?

a) Não é evidente que um programa amplamente utilizado tenha impacto?

- Uma das motivações da avaliação de impacto é, precisamente, a validação das teorias que serviram de base ao desenho do programa.
- Por esse motivo, é tão comum e importante a avaliação de impacto de programas em escala-piloto. A avaliação de um piloto busca verificar se a ligação entre o acesso ao programa e os resultados esperados é válida, para que o programa possa ser expandido com menos incerteza sobre o seu potencial de alcançar sucesso.
- Similarmente, uma avaliação de impacto também é útil para confirmar se as percepções dos usuários sobre a eficácia do programa coincidem com a realidade.

Por que avaliar o impacto de um programa social?

a) Não é evidente que um programa amplamente utilizado tenha impacto?

- Por outro lado, podemos argumentar que o objetivo da avaliação de impacto vai muito além de simplesmente determinar a existência ou não de impacto.
- Mesmo que tenhamos convicção da existência de impacto de dado programa, ainda assim será necessário conduzir a avaliação para que possamos “entender” o impacto.
- Tanto para o gestor do programa, quanto para os usuários, tão ou mais importante do que saber se há impacto é conhecer a magnitude dele, assim como saber sobre quais dimensões o programa tem impacto e que canais permitem que ele se manifeste.
- Esse conhecimento é indispensável para o aperfeiçoamento do desenho do programa e para a melhoria da sua adequação às necessidades de seus usuários.

Por que avaliar o impacto de um programa social?

b) Demanda pelos mais ricos é evidência de impacto?

- Muitos dos programas sociais existentes são ofertados gratuitamente para a parcela mais pobre da sociedade.
- Nesse caso, a adesão voluntária ao programa deve ser necessariamente interpretada como indício de impacto do programa? Pode-se argumentar que não.
- É possível que os pobres tenham conhecimento limitado sobre a eficácia daquele serviço ou programa. **Além disso, como a oferta é gratuita, os beneficiários aderem porque não perdem nada participando.**

Por que avaliar o impacto de um programa social?

b) Demanda pelos mais ricos é evidência de impacto?

- Contudo, o que dizer do impacto de um serviço oferecido gratuitamente aos mais pobres, mas para o qual existe provisão privada para suprir a demanda dos mais ricos, que pagam preços significativos pelo acesso?
- Assumindo que a qualidade dos serviços oferecidos gratuitamente aos pobres seja igual à qualidade da oferta privada aos ricos, não seria a existência de demanda pelos mais ricos evidência suficiente de que o serviço teria impacto?
- Com base na teoria econômica, podemos argumentar que a existência de demanda por um serviço com preço significativo é sim o indício de impacto. Por que então a necessidade de realizar uma avaliação de impacto nessa situação?

Por que avaliar o impacto de um programa social?

b) Demanda pelos mais ricos é evidência de impacto?

- Em primeiro lugar, observadas as diferenças entre os pobres e os ricos, pode-se argumentar que o fato de um serviço ou programa ter impacto sobre um grupo não implica necessariamente que terá o mesmo impacto sobre o outro grupo.
- Dessa forma, havendo demanda dos mais ricos por um serviço, ainda assim será necessário avaliar a magnitude do impacto sobre os mais pobres.

Por que avaliar o impacto de um programa social?

b) Demanda pelos mais ricos é evidência de impacto?

- Em segundo lugar, mesmo quando não há dúvidas sobre a existência de impacto do serviço, pode ser importante avaliar a sua magnitude sobre os mais pobres. Sabemos que, em geral, as prioridades dos indivíduos (e, portanto, o valor dos benefícios) variam com o nível de renda.
- Benefícios que têm elevado valor para os mais ricos podem não ser classificados como prioritários para os mais pobres.
- Logo, a evidência da eficácia do programa entre os mais ricos pode não ser suficiente para justificar a implantação de um programa análogo para os mais pobres. Nesses casos, justifica-se, pois, a utilização da avaliação de impacto para produzir evidência direta do impacto do programa e do valor desses benefícios para o grupo dos mais pobres.

Por que avaliar o impacto de um programa social?

c) A relação entre impacto, ambiente socioeconômico e características da população beneficiada

- Formuladores de políticas públicas, geralmente, desenham um programa social visando a uma população-alvo específica sendo beneficiada em um ambiente preestabelecido.
- Assim, é natural que existam poucas dúvidas sobre o impacto de determinado programa quando esse é implementado no local e momento corretos e beneficia a população para a qual foi originalmente desenhado.
- **Porém, esse mesmo programa implementado em um ambiente completamente distinto do planejado provavelmente não terá o mesmo resultado.** É de se esperar que a mesma ação dirigida a outra população, outro local e outro momento do tempo não beneficie da mesma forma os que dela participam.

Por que avaliar o impacto de um programa social?

c) A relação entre impacto, ambiente socioeconômico e características da população beneficiada

- Uma das utilidades da avaliação de impacto é, precisamente, determinar em que medida a eficácia de um programa depende das características da população beneficiada, do momento no tempo e da natureza do ambiente socioeconômico em que ocorre.

Por que avaliar o impacto de um programa social?

d) Impacto potencial versus impacto efetivo

- O impacto de um programa não depende apenas do seu desenho e da sua adequação ao perfil dos beneficiários e ao ambiente socioeconômico em que esses vivem.
- **Igualmente determinante da magnitude do impacto é a forma como o programa é implementado.** Mesmo sendo poucas as dúvidas sobre o impacto potencial de um programa bem desenhado e focalizado, podem persistir grandes incertezas relativas ao efetivo impacto desse programa quando implementado de maneira deficiente.
- **Nesse caso, o objetivo da avaliação de impacto não é investigar o impacto teórico ou potencial do programa, mas sim seu impacto efetivo, dada a maneira imperfeita como foi implementado.**

Por que avaliar o impacto de um programa social?

d) Impacto potencial versus impacto efetivo

- Em suma, o impacto de um programa depende não apenas do seu impacto potencial, mas está também intrinsecamente ligado às condições da sua implementação.
- Ainda que um programa tenha comprovadamente impacto potencial, é fundamental avaliar o seu impacto real. Nesse caso, a avaliação de impacto é útil para verificar qual parcela do impacto potencial foi desperdiçada devido a falhas no processo de implementação.

Por que avaliar o impacto de um programa social?

d) Impacto potencial versus impacto efetivo

- Esse dilema é inerente à avaliação de impacto de políticas públicas. Em várias áreas, existem estudos teóricos que, consensualmente, apontam para a importância de determinadas ações. Em contraste, muitas avaliações encontram impactos irrisórios dessas mesmas ações após a implementação.
- Assim, o frequente descompasso entre os resultados teóricos e empíricos reforça ainda mais a necessidade de uma avaliação de impacto, essencial para discernir entre o impacto do programa como originalmente desenhado (impacto potencial) e o impacto do programa como de fato implementado (impacto efetivo).

Por que avaliar o impacto de um programa social?

e) Avaliação de impacto como forma de monitoramento

- **Os impactos de determinado programa social dificilmente são invariáveis ao longo do tempo. Assim, nada garante que um programa avaliado hoje continue tendo o mesmo impacto no próximo ano ou no ano subsequente.**

Por que avaliar o impacto de um programa social?

e) Avaliação de impacto como forma de monitoramento

- Podemos apontar pelo menos quatro razões para justificar essa potencial inconstância temporal do impacto.
- **Primeiramente, o impacto de um programa, normalmente, está relacionado com o tempo de exposição do beneficiário, podendo tanto declinar rapidamente logo após o término do programa como persistir por longo prazo.**
- **Em segundo lugar, com o tempo podem ocorrer flutuações na qualidade da gestão do programa e, conseqüentemente, na eficácia da implementação dele, gerando oscilações na magnitude do impacto.**

Por que avaliar o impacto de um programa social?

e) Avaliação de impacto como forma de monitoramento

- **Em terceiro lugar, o impacto pode variar ao longo do tempo se for sensível a mudanças no ambiente socioeconômico.**
- **Por fim, se o perfil da população beneficiada pelo programa se modifica ao longo do tempo, é de se esperar que a magnitude do impacto também se altere.**

Por que avaliar o impacto de um programa social?

e) Avaliação de impacto como forma de monitoramento

- Devido a cada um desses motivos, mesmo programas que têm impacto inicial de magnitude conhecida devem ser continuamente avaliados.
- Por um lado, uma avaliação contínua é a única maneira de identificar os impactos de médio e longo prazos do programa e, portanto, a única forma de avaliar seu grau de maturação e persistência.
- Por outro lado, repetidas avaliações do mesmo programa podem revelar como o impacto do programa responde a mudanças no ambiente socioeconômico e no perfil dos beneficiários.

Por que avaliar o impacto de um programa social?

e) Avaliação de impacto como forma de monitoramento

- Por fim, mesmo no caso em que o perfil da população beneficiada e o ambiente socioeconômico permanecem estáveis, avaliações de impacto periódicas podem ser de extrema utilidade como instrumento de gestão.
- Nesse caso, flutuações na magnitude do impacto podem indicar melhoria ou deterioração na forma de funcionamento e, portanto, na gestão do programa.

Por que avaliar o impacto de um programa social?

e) Avaliação de impacto como forma de monitoramento

- Os argumentos apresentados sinalizam que o impacto de um programa tem caráter mutável ao longo do tempo.
- Portanto, a avaliação de impacto deve ser considerada como ferramenta de utilização contínua, já que as conclusões auferidas a partir de sua utilização pontual não são necessariamente válidas intertemporalmente.

Por que avaliar o impacto de um programa social?

f) A questão do custo-efetividade

- A existência de impacto não é suficiente para justificar a alocação de recursos adicionais para dado programa.
- Em uma sociedade, se existirem programas alternativos que perseguem os mesmos objetivos, a opção por determinado programa dependerá de ele produzir o maior impacto por unidade de custo. Esse critério é comumente denominado custo-efetividade.
- Assim, em um cenário onde os recursos são escassos e diferentes programas competem pelos mesmos recursos, a avaliação de impacto é essencial para mensurar a magnitude do impacto e calcular a relação custo-efetividade de cada programa.

Por que avaliar o impacto de um programa social?

f) A questão do custo-efetividade

- Mesmo que o impacto de dado programa seja inquestionável, esse programa pode não ser aquele com a melhor relação custo-efetividade e, portanto, pode não ser o melhor candidato a receber os recursos disponíveis.
- Assim, para se avaliar a relação custo-efetividade de um programa, é necessário não apenas reconhecer a existência do impacto, mas também mensurar a sua magnitude.

Por que avaliar o impacto de um programa social?

g) A questão do custo-benefício

- Para que recursos sejam alocados a um programa, não basta que tenha impacto ou mesmo que seja aquele com a melhor relação custo-efetividade.
- É necessário também que o valor dos benefícios do programa supere seus custos, isto é, que o programa tenha uma relação custo-benefício favorável.
- No cálculo dessa relação, deve-se considerar todos os custos e benefícios sociais, tanto aqueles sobre os agentes diretamente envolvidos como as externalidades sobre o restante da sociedade.

Por que avaliar o impacto de um programa social?

h) Por que não investigar diretamente a propensão a pagar?

- A relação custo benefício de uma política pública ou programa social compara os custos e o valor dos benefícios do programa, sendo o último obtido a partir do produto entre a magnitude do impacto e o valor atribuído a ele pelo conjunto de beneficiários.

Por que avaliar o impacto de um programa social?

h) Por que não investigar diretamente a propensão a pagar?

- Porém, a avaliação de impacto apresenta duas limitações.
- Primeiramente, um programa tem geralmente variedade de impactos, alguns intencionais e muitos outros colaterais. Nesse caso, para obter a relação custo-benefício, seria necessário estimar a magnitude de todos esses impactos, o que já seria um desafio, dado que muitos desses impactos são de difícil mensuração.
- Em segundo lugar, mesmo que fosse possível estimar a magnitude de todos os impactos, ainda seria preciso estimar o valor que cada beneficiário atribui a cada um deles. De fato, a avaliação de impacto é apenas o primeiro passo para se estimar a relação custo-benefício.

Por que avaliar o impacto de um programa social?

h) Por que não investigar diretamente a propensão a pagar?

- No entanto, existe uma alternativa para estimar o valor dos benefícios: a “propensão a pagar” do beneficiário pelo serviço.
- O valor total de um programa, benefício ou serviço para o beneficiário pode ser avaliado a partir da investigação do valor pelo qual ele estaria disposto a trocar o acesso ao programa, medido em termos monetários ou em termos do acesso a outros bens e serviços.

Por que avaliar o impacto de um programa social?

h) Por que não investigar diretamente a propensão a pagar?

- As técnicas utilizadas para a investigação da “propensão a pagar” podem ser classificadas em dois grandes grupos: (i) comportamental e (ii) não comportamental.

Por que avaliar o impacto de um programa social?

h) Por que não investigar diretamente a propensão a pagar?

- Nos métodos não comportamentais, pergunta-se diretamente ao beneficiário quanto ele estaria disposto a pagar pelo acesso a determinado programa, benefício ou serviço.
- O problema, nesse caso, é que frequentemente a resposta pode não ser fidedigna.
- Como a pergunta se refere a uma situação hipotética, as pessoas tendem a superestimar o valor que elas efetivamente estariam dispostas a pagar. Existem, entretanto, protocolos preestabelecidos que definem como investigações dessa natureza podem ser conduzidas, visando mitigar a possibilidade de resultados pouco confiáveis.

Por que avaliar o impacto de um programa social?

h) Por que não investigar diretamente a propensão a pagar?

- Por outro lado, nos métodos comportamentais, o valor do benefício é inferido a partir da observação direta do efetivo comportamento do beneficiário.
- No caso de serviços para os quais exista mercado, verificamos que o preço de mercado é o valor (ou a propensão a pagar) que o beneficiário marginal atribui ao serviço em questão. Quando não existe mercado, ainda assim o valor associado ao serviço pode ser inferido com base na observação de outros comportamentos.
- No caso de um posto de saúde, por exemplo, o valor pode ser avaliado a partir do raio de cobertura do serviço. Quanto mais valorizado o serviço, maior a distância que os beneficiários estarão dispostos a percorrer para acessá-lo.

Por que avaliar o impacto de um programa social?

h) Por que não investigar diretamente a propensão a pagar?

- Em resumo, a vantagem metodológica da utilização da propensão a pagar é que esse método estima diretamente o valor do benefício, evitando problemas potenciais causados pela necessidade de estimar múltiplos impactos e o valor atribuído a cada um deles.

Por que avaliar o impacto de um programa social?

h) Por que não investigar diretamente a propensão a pagar?

- Ainda assim, existem essencialmente três justificativas que sustentam o uso da avaliação de impacto.
- Em primeiro lugar, quando o impacto de um programa ocorre por meio de poucas dimensões de fácil mensuração, a magnitude do impacto pode ser avaliada de forma relativamente incontestável com base em métodos experimentais.
- Além disso, na ausência de mercados para o serviço ou benefício em questão, a utilização da propensão a pagar poderia levar a estimativas ainda menos confiáveis. No entanto, é importante ressaltar que o aumento do número de dimensões do impacto e o surgimento de mercado para o serviço ou benefício em questão tornaria o uso da propensão a pagar mais fundamentada *vis a vis* à utilização da avaliação de impacto.

Por que avaliar o impacto de um programa social?

h) Por que não investigar diretamente a propensão a pagar?

- Em segundo lugar, pode-se argumentar que a utilização da propensão a pagar tem sua confiabilidade reduzida quando existem externalidades, e o programa em questão tem impactos sobre não beneficiários.
- Nesse caso, se o número de dimensões do impacto é limitado e o grupo de não beneficiários impactados pelo programa está bem definido, uma avaliação de impacto experimental é provavelmente capaz de produzir estimativas mais confiáveis.

Por que avaliar o impacto de um programa social?

h) Por que não investigar diretamente a propensão a pagar?

- Por fim, o uso da propensão a pagar tem a grande limitação de não permitir a identificação dos motivos pelos quais os beneficiários valorizam o serviço.
- Assim, sabe-se quanto o programa é valorizado, mas nada se sabe sobre os mecanismos pelos quais ele influencia o bem-estar dos beneficiários.

Por que avaliar o impacto de um programa social?

h) Por que não investigar diretamente a propensão a pagar?

- Essa limitação é especialmente problemática no caso de programas multifacetados, à medida que é impossível distinguir qual componente do programa está contribuindo mais ou menos para a satisfação do usuário.
- Do ponto de vista dos beneficiários, é possível que seja suficiente saber o grau de satisfação dos beneficiários com o programa de modo geral. Porém, para os gestores e aqueles que financiam o programa, essa informação pode ser insuficiente.

Por que avaliar o impacto de um programa social?

h) Por que não investigar diretamente a propensão a pagar?

- Existem duas razões que justificam a necessidade de identificar os impactos específicos de um programa.
- Por um lado, é possível que os responsáveis pelo financiamento valorizem os impactos específicos de maneira distinta dos beneficiários.
- Em outras palavras, aqueles que financiam o programa podem querer se assegurar de que ele esteja causando satisfação pela via pretendida e não por outro motivo qualquer.

Por que avaliar o impacto de um programa social?

h) Por que não investigar diretamente a propensão a pagar?

- Por outro lado, identificar a magnitude dos impactos específicos é importante para avaliar a adequação do programa no alcance dos seus objetivos específicos.
- Essa informação pode ser útil tanto para o redesenho do próprio programa quanto para o desenho de outros programas em contextos dos quais alguns impactos específicos sejam particularmente valorizados.

Por que avaliar o impacto de um programa social?

i) Heterogeneidade do programa e da população alvo

- A grande maioria das intervenções é heterogênea e oferecida a uma clientela também heterogênea.
- Na maioria dos casos, é possível ajustar os diversos parâmetros do programa, desde sua duração e intensidade até seu conteúdo e qualidade.
- É de se esperar, portanto, que a magnitude do impacto varie com esses parâmetros e, também, segundo as características do beneficiário e o contexto socioeconômico em que o programa se insere.

Por que avaliar o impacto de um programa social?

i) Heterogeneidade do programa e da população alvo

- Assim, a realização de uma avaliação de impacto é essencial para entender como a heterogeneidade na intervenção e no perfil dos usuários se reflete nos resultados alcançados.
- O mesmo programa pode ter muitos desenhos que variam em termos de eficiência. A avaliação de impacto possibilita não apenas identificar o desenho mais eficiente, mas também identificar os segmentos da população-alvo para os quais os benefícios são maiores.

Por que avaliar o impacto de um programa social?

i) Heterogeneidade do programa e da população alvo

- A identificação desses grupos é fundamental para o desenho do programa. Em curto prazo, a atuação do programa deveria se concentrar nos segmentos da população-alvo que têm maior probabilidade de se beneficiar dessas ações.
- Por outro lado, a identificação dos segmentos que pouco se beneficiam das ações é igualmente importante à medida que revela a necessidade do investimento em ações alternativas voltadas para esses grupos.

Quando avaliar o impacto de um programa social?

- Existem essencialmente três momentos na execução de um programa em que avaliações de impacto podem ocorrer: **(i) antes do início do programa (são as ditas avaliações *ex ante*), (ii) durante o período de execução do programa (são as ditas avaliações *ex post* de percurso) e (iii) após a conclusão do programa (são as ditas avaliações *ex post* de encerramento).**
- Em cada caso, a avaliação é caracterizada por metodologias e objetivos distintos, condicionados à disponibilidade de informações relacionadas ao programa naquele momento do tempo.

Quando avaliar o impacto de um programa social?

a) Avaliação *ex ante* e *ex post*

- A decisão pela implementação de dado programa é geralmente precedida pela apresentação de argumentos que justificam a racionalidade da opção por aquela determinada intervenção.
- **Prever a magnitude dos impactos antes que a intervenção seja implantada é o objetivo das avaliações *ex ante* de impacto.**
- **As avaliações *ex ante* buscam estimar os custos e os impactos esperados do programa em questão.**

Quando avaliar o impacto de um programa social?

a) Avaliação *ex ante* e *ex post*

- **Avaliações de impacto realizadas durante o período de operação do programa, denominadas avaliações *ex post* de percurso, são utilizadas para verificar a validade das previsões realizadas pelas avaliações *ex ante*.**
- Dessa forma, esse tipo de avaliação mostra-se fundamental para traçar recomendações sobre o futuro do programa, no sentido de interromper, continuar ou aperfeiçoar as ações da intervenção em curso.
- Esse tipo de avaliação é muito comum durante a fase piloto da implementação de um programa, embora seja importante que a sua aplicação não se limite a essa fase.

Quando avaliar o impacto de um programa social?

a) Avaliação *ex ante* e *ex post*

- O grande desafio para a realização de uma boa avaliação *ex post* de percurso é o tempo.
- Por um lado, é importante que a avaliação seja realizada quanto antes para que tenha sua influência sobre o desenho, operação, continuidade e expansão do programa maximizada.
- Por outro lado, dado que alguns impactos podem levar tempo para se manifestar, quanto mais cedo realizada a avaliação de impacto, maior a probabilidade de se subestimar ou até ignorar impactos importantes que não se manifestaram no curto prazo.

Quando avaliar o impacto de um programa social?

a) Avaliação *ex ante* e *ex post*

- Esse problema é particularmente grave já que, frequentemente, os impactos de maior importância podem ser os de longo prazo.
- Assim, toda avaliação *ex post* de percurso precisa analisar cuidadosamente o *trade-off* entre fornecer respostas rápidas para auxiliar no aperfeiçoamento e expansão do programa e aguardar o tempo necessário para que parte substancial dos impactos do programa se manifeste.
- Portanto, avaliações *ex post* de percurso são mais comuns em programas de longa duração ou mesmo de duração ilimitada, sendo difíceis de serem realizadas em programas de curta duração.

Quando avaliar o impacto de um programa social?

a) Avaliação *ex ante* e *ex post*

- Ou seja, muitas vezes é necessário uma maturação adequada do programa para que se possa medir os impactos durante a operação da política.

Quando avaliar o impacto de um programa social?

a) Avaliação *ex ante* e *ex post*

- A realização de uma avaliação de impacto também é justificada após o encerramento de um programa social.
- Nesse caso, dois objetivos merecem atenção. Por um lado, a avaliação *ex post* de encerramento é útil para determinar quão adequada foi a decisão de alocação de recursos àquele programa.
- Trata-se de uma avaliação *ex post* das decisões realizadas no passado e da qualidade da gestão do programa. Não permite rever ou aprimorar ações, mas apenas verificar se os agentes envolvidos no processo decisório tomaram ou não decisões adequadas e se os responsáveis pelo programa fizeram uma boa gestão.

Quando avaliar o impacto de um programa social?

a) Avaliação *ex ante* e *ex post*

- Por outro lado, uma utilidade extremamente importante das avaliações *ex post* em geral está relacionada à identificação de melhores práticas.
- Avaliações *ex post* de programas com objetivos similares geralmente fornecem informações relevantes que podem auxiliar na tomada de decisão de gestores públicos empenhados em desenhar novos programas ou aperfeiçoar os programas existentes na área em questão.

Quando avaliar o impacto de um programa social?

a) Avaliação *ex ante* e *ex post*

- **Outra vantagem da avaliação *ex post* é que não há restrição de tempo para a sua realização, o que torna esse tipo de avaliação ideal para garantir a captação dos impactos verdadeiramente de longo prazo de um programa social.**

Quando avaliar o impacto de um programa social?

b) Diferenças informacionais e metodológicas

- Em decorrência do momento em que são realizadas, as avaliações *ex ante* e *ex post* se baseiam cada qual em um conjunto distinto de informações, aquelas disponíveis naquele dado momento do tempo.
- Avaliações *ex post* podem contar com informações sobre a situação de beneficiários e não beneficiários antes e em vários momentos após o início da intervenção.
- Por outro lado, avaliações *ex ante* são conduzidas às vezes sem informações sobre quem serão os efetivos beneficiários do programa. À medida que informações naturalmente se acumulam ao longo do tempo, avaliações *ex post* têm grande vantagem informacional sobre suas contrapartidas *ex ante*.

Quando avaliar o impacto de um programa social?

b) Diferenças informacionais e metodológicas

- Na avaliação *ex ante*, o papel da teoria passa a ser primordial, uma vez que as estimativas de impacto são obtidas a partir de simulações com dados na maioria das vezes secundários sobre o comportamento hipotético dos futuros beneficiários do programa a ser implantado.
- Nesse caso, é com base na teoria sobre o comportamento desses beneficiários que são derivadas as hipóteses a partir das quais essas simulações de impacto são realizadas.

Modelo de Resultados Potenciais

Introdução

- O problema central da área de avaliação de programas é construir o contrafactual, ou contrafactual, do grupo tratado pelo programa.
- Na forma mais simples de apresentar esse problema, podemos pensar que qualquer indivíduo está sempre em uma de duas situações distintas: ter sido ou não tratado pelo programa.
- Idealmente, o melhor grupo de comparação para os indivíduos tratados seria formado pelos mesmos indivíduos na situação em que eles não fossem tratados.
- Contudo, essas situações são mutuamente exclusivas: claramente não é possível observar os mesmos indivíduos na condição de tratados e de não tratados ao mesmo tempo.

Introdução

- O desafio do avaliador, portanto, é encontrar um grupo de indivíduos que represente adequadamente a situação de não tratamento, ou seja, um grupo que funcione como bom contrafactual do grupo tratado.

Introdução

- Uma série de pontos precisam ser levados em consideração para se encontrar um grupo de controle que possa replicar o grupo tratado, caso ele não tivesse passado pela intervenção.
- Naturalmente, como não é possível observar essa situação contrafactual, a escolha desse grupo de comparação envolverá o uso de procedimentos e hipóteses, cuja finalidade básica é minimizar o que se costuma denominar de viés de autosseleção.

Introdução

- Para entender melhor o que está por trás desse viés, apresentaremos a ideia mais utilizada na área de avaliação, a saber, a ideia de resultados potenciais.
- Elemento crucial para bom entendimento dos resultados potenciais – e da área de avaliação de impacto em geral – é ter sempre em mente a potencial existência de uma situação contrafactual.

Introdução

- Antes de apresentarmos os resultados potenciais, vale a pena analisar dois métodos “ingênuos”, porém amplamente utilizados por não especialistas para construir o grupo de controle.
- O primeiro é baseado na comparação do que ocorre com o grupo de tratamento antes e depois da intervenção.
- O segundo consiste em comparar o grupo que escolhe passar pelo programa com um grupo que decide não participar da intervenção. Como veremos, esses métodos dificilmente são capazes de fornecer o impacto do programa.

Comparando os tratados antes e depois

- Considere a situação em que um programa de treinamento profissional foi oferecido para trabalhadores de baixa escolaridade em certa localidade, por exemplo, um município ou uma área de uma cidade.
- Esse programa consistia de um curso ministrado por pessoal especializado e tinha como objetivo aumentar o rendimento do trabalho dos participantes.
- Suponha que o total de inscritos tenha sido igual ao número de vagas oferecidas e que todos os inscritos tenham participado do programa.

Comparando os tratados antes e depois

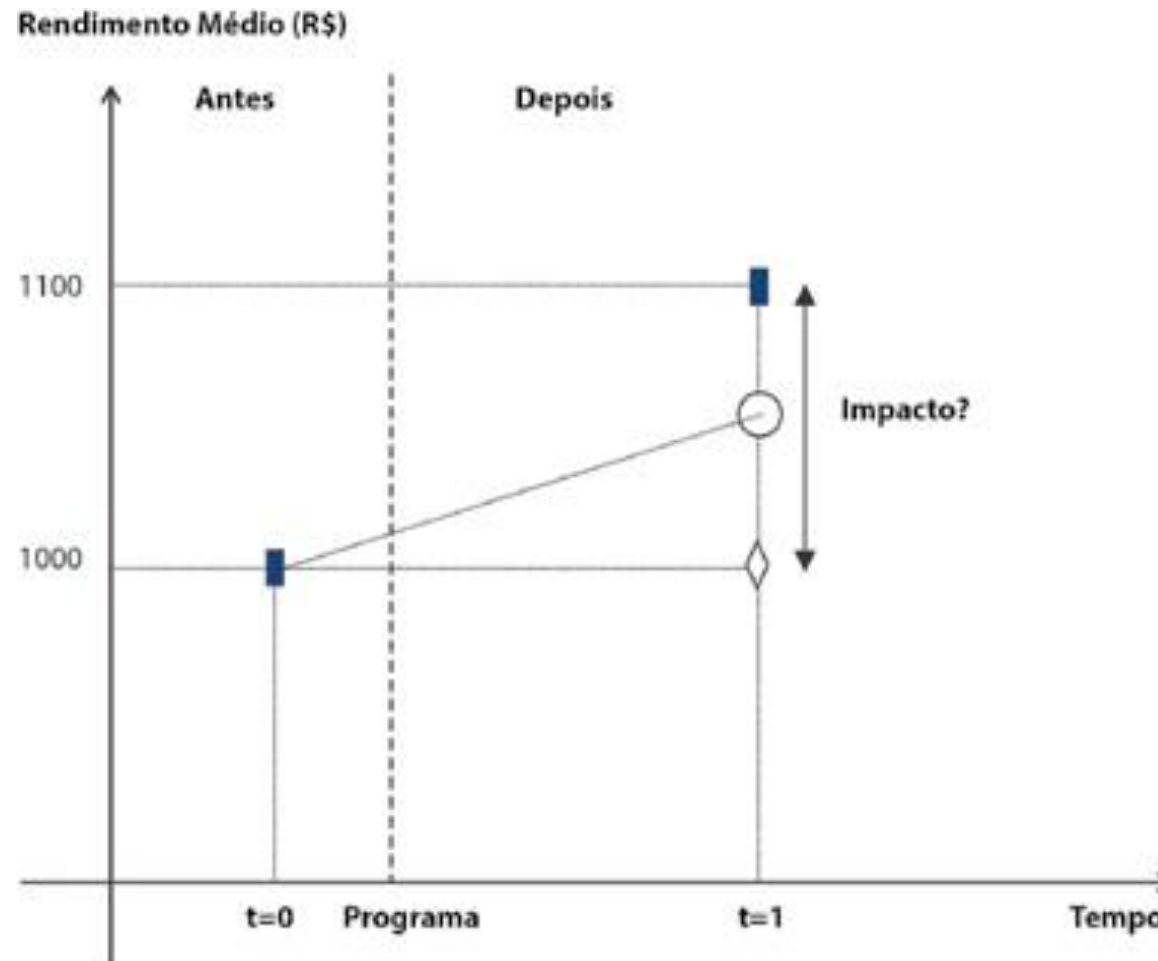
- Pouco antes do começo do programa (por exemplo, no dia da inscrição), todos os participantes responderam a um questionário que levantou informações sobre suas características socioeconômicas, incluindo a situação de emprego e rendimento laboral naquele momento.
- Suponha que o rendimento médio do trabalho para todos os indivíduos que entraram no programa de treinamento tenha sido de R\$ 1.000.

Comparando os tratados antes e depois

- Um ano depois, os mesmos tipos de informações foram novamente coletadas com todos os participantes do programa. O rendimento médio encontrado nesse segundo momento foi igual a R\$ 1.100.
- A Figura 2.1 ilustra as informações obtidas para o rendimento médio para os períodos anterior e posterior ao programa. Os retângulos estão associados aos valores observados.
- Alguns diriam que o efeito do programa foi de R\$ 100 (R\$ 1.100 – R\$ 1.000), o que equivale ao aumento de 10% no rendimento médio inicial. Mas será que podemos afirmar que esse foi o verdadeiro efeito do programa?

Comparando os tratados antes e depois

Figura 2.1: Rendimento médio dos participantes antes e depois do programa



Comparando os tratados antes e depois

- Para responder a essa pergunta, temos que nos questionar se o grupo de tratamento antes do programa representa o contrafactual adequado para a situação de não tratamento depois do programa.
- Ou seja, em termos da Figura 2.1, será que o losango representa o valor que seria recebido pelo grupo tratado depois do programa, caso ele não tivesse passado pela intervenção?
- A resposta a essa pergunta seria normalmente negativa, já que há uma série de fatores que podem ter afetado o rendimento médio do grupo tratado além do programa.

Comparando os tratados antes e depois

- Por exemplo, é possível que a situação do mercado de trabalho da localidade onde residem os participantes do programa tenha melhorado ao longo dos doze meses entre as coletas das informações.
- Para ilustrar essa possibilidade, suponha que essa mudança nas condições do mercado de trabalho tenha sido responsável por levar o rendimento médio para o ponto associado à forma circular na Figura 2.1.
- Nesse caso, a variação do rendimento médio que pode ser atribuída ao programa deixa de ser os R\$ 100 estimados inicialmente.

Comparando os tratados antes e depois

- Esse exemplo mostra que, a menos que possamos controlar a influência do mercado de trabalho e de todos os outros fatores que afetam o rendimento médio do trabalho, a simples comparação entre o que ocorre com o grupo de tratados antes e depois do programa não identifica corretamente o efeito causal de uma intervenção.

Comparando os tratados com os não tratados

- Outra prática “ingênua” comumente usada pelos não especialistas é a que compara o grupo de tratados diretamente com o grupo de não tratados.
- Entretanto, apenas por não ter passado pela intervenção não significa que o grupo de não tratados representa bem o que ocorreria com o grupo de tratamento, caso este não tivesse sido tratado.
- Em outras palavras, a ausência do tratamento para alguns não gera automaticamente o contrafactual de não tratamento para outros.

Comparando os tratados com os não tratados

- **As principais razões para isso estão relacionadas com as diferenças nas características observáveis e não observáveis (pelo avaliador) entre os dois grupos.**

Comparando os tratados com os não tratados

- As principais razões para isso estão relacionadas com as diferenças nas características observáveis e não observáveis (pelo avaliador) entre os dois grupos.
- **Por exemplo, é possível que, mesmo que os grupos sejam parecidos (em média) em atributos observáveis, como gênero, idade, escolaridade, renda familiar, etc, eles difiram em habilidades, motivação, esforço e outras características que tipicamente não estão disponíveis nas bases de dados utilizadas nas avaliações de impacto.**
- Na medida em que muitas dessas características possam afetar a variável de resultado sobre a qual se quer medir o impacto da intervenção, o uso da comparação direta entre os dois grupos não, necessariamente, mede o efeito causal do programa. **Afinal, as diferenças nas características não observáveis entre os grupos poderiam estar misturadas ao efeito do programa.**

Comparando os tratados com os não tratados

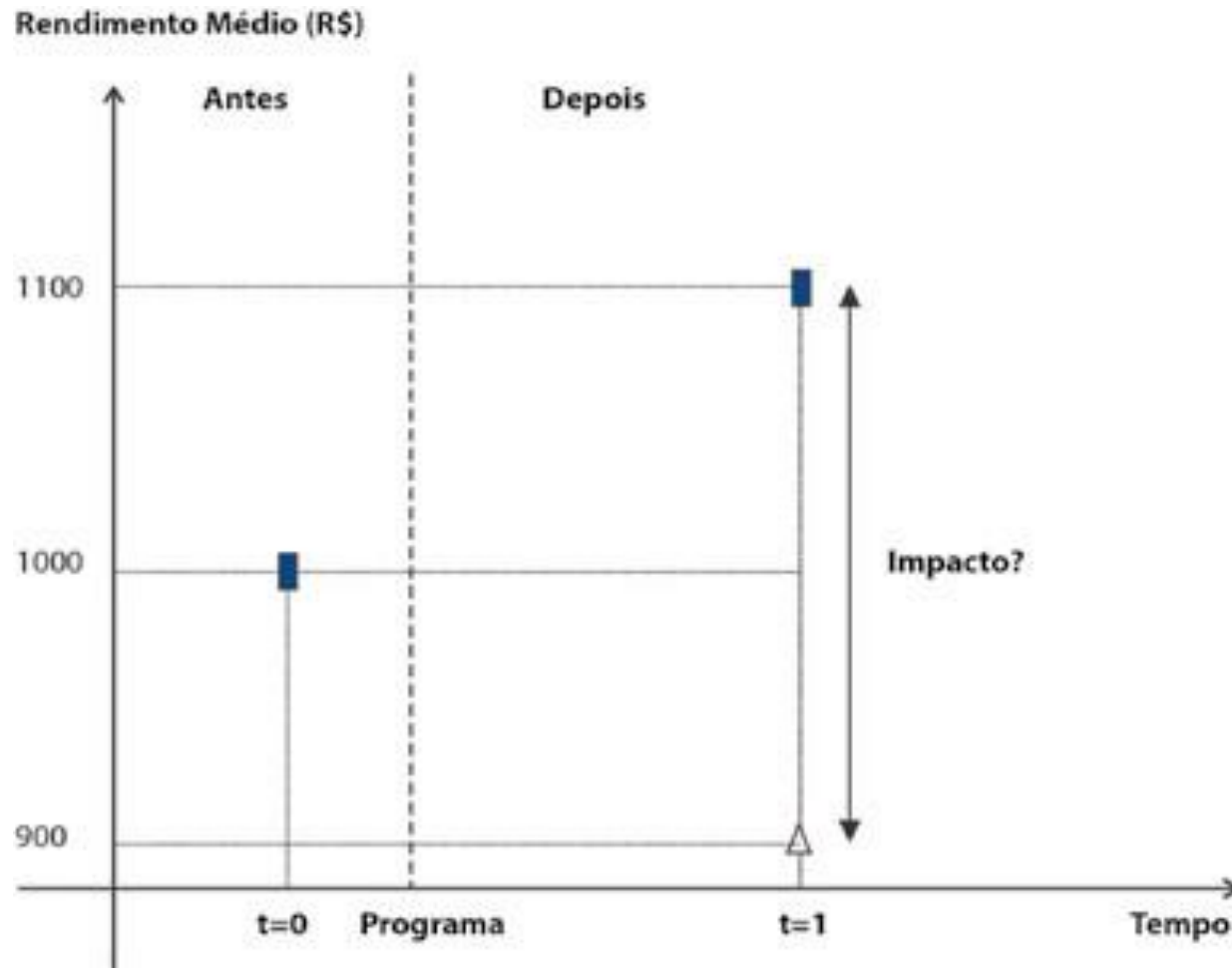
- Para ilustrar esse problema, continuaremos com a situação hipotética da subseção anterior.
- Assim, admita que o segundo questionário usado com os participantes do programa de treinamento tenha sido aplicado a uma amostra de pessoas da mesma localidade que não tenham participado do programa.
- Suponha que a média do rendimento do trabalho obtida com as informações coletadas com esse grupo de não tratados após o programa tenha sido de R\$ 900.

Comparando os tratados com os não tratados

- Suponha que os dois grupos sejam idênticos, em média, nas características levantadas no questionário, que não incluiu nenhuma tentativa de medir a motivação das pessoas.
- A Figura 2.2 contém as informações para os dois grupos para o período depois do programa, onde o triângulo representa o rendimento médio do grupo de não participantes.

Comparando os tratados com os não tratados

Figura 2.2: Rendimento médio dos tratados e não tratados



Comparando os tratados com os não tratados

- Embora as pessoas dos dois grupos sejam parecidas na média em diversas características, não podemos afirmar que os R\$ 200 de diferença no rendimento médio dos grupos (ou seja, a distância entre o retângulo e o triângulo em $t=1$) medem corretamente o impacto do programa de treinamento.
- Para isso, basta admitirmos que a motivação (que não foi medida) seja uma característica que afeta o desempenho das pessoas no mercado de trabalho, incluindo o seu salário.
- Se os dois grupos diferem em termos de motivação – diferença essa que pode explicar por que o grupo tratado decidiu participar do programa, e o grupo não tratado preferiu não participar –, então é provável que parte dos R\$ 200 de diferença entre os grupos se deva ao fator motivação. Se isso ocorrer, não podemos tratar essa quantia como o efeito causal do programa de treinamento.

Comparando os tratados com os não tratados

- Há um conjunto amplo de metodologias que buscam isolar o efeito dos programas dos efeitos causados por outros fatores que afetam a(s) variável(eis) de resultado de interesse.
- **A ideia então é tentar construir um verdadeiro grupo de controle, um grupo de controle construído com base nas hipóteses adequadas e nas melhores práticas para a mensuração do impacto.**

Comparando os tratados com os não tratados

- Podemos resumir os problemas da comparação ingênua entre tratados e não tratados da seguinte forma:
- Por um lado é possível medir, em média, as características observáveis dos indivíduos, sejam aqueles que pertencem ao grupo de tratamento ou ao grupo de não tratados.
- Porém, as características não observáveis são o grande problema.
- Ou seja, problema não está nas características observáveis, mas sim nas não observáveis.

Comparando os tratados com os não tratados

- **Ainda que as características observáveis do grupo de tratados e não tratados sejam, em média, bem parecidas, as características não observáveis geram uma dificuldade natural para a comparação.**

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

- Os métodos de avaliação de impacto buscam solucionar o problema do viés de autosseleção.
- Cada método lança mão de certos procedimentos e hipóteses para tentar resolver esse problema.

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

a) Aleatorização

- Considerado o padrão-ouro da área de avaliação, o primeiro método baseia-se na **aleatorização** de indivíduos, famílias ou até mesmo de comunidades inteiras para participar ou não do programa.
- Esse procedimento de aleatorização gera dois grupos experimentais, o de tratamento, formado pelos indivíduos que foram aleatorizados para participar do programa, e o grupo de controle, composto pelos que foram aleatorizados para ficar de fora

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

a) Aleatorização

- **O fato de que a participação ou não no tratamento é definida pelo procedimento de aleatorização garante que os grupos de tratamento e controle sejam parecidos tanto nas características observáveis quanto nas não observáveis.**
- O viés de autosseleção fica assim contornado, permitindo que a comparação entre os grupos identifique o efeito causal do programa.

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

a) Diferenças em Diferenças (DID)

- O método de Diferenças em Diferenças (DID) requer **dados no painel** das unidades tratadas e não tratadas, com períodos de observação antes e depois da intervenção.
- Ele recebe esse nome porque **duas diferenças são calculadas: a diferença entre os períodos de tempo para cada unidade observada e a diferença entre as unidades para cada período de tempo.**
- Como os dados são em painel, o método comporta a presença de um termo invariante no tempo para cada unidade de observação.

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

a) Diferenças em Diferenças (DID)

- **Esse termo, que é eliminado no cálculo da diferença temporal para cada unidade, capta todas as características não observadas da unidade que são constantes no tempo.**
- **A principal hipótese desse método é que a trajetória da variável de resultado do grupo de controle reflita a do grupo de tratamento na ausência da intervenção.**

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

a) Diferenças em Diferenças (DID)

- Embora essa hipótese não seja testável, uma indicação de sua validade ocorre quando os grupos possuíam a mesma trajetória temporal da variável de resultado antes da intervenção.
- A ideia é que, se as trajetórias dos grupos são semelhantes no período pré-programa, então aumentam as chances de que a trajetória do grupo de controle após o programa represente bem a trajetória do grupo de tratamento na ausência da intervenção.

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

a) Diferenças em Diferenças (DID)

- Nesse caso, qualquer diferença entre os grupos que apareça após o programa pode ser interpretada como o impacto da intervenção.
- Quando as informações disponíveis cobrem um intervalo relativamente extenso antes do programa, o teste de trajetórias iguais entre os grupos pré-programa é passível de ser testada estatisticamente nos dados.

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

a) Diferenças em Diferenças (DID)

- Em resumo, o método DID baseia-se no cálculo de uma dupla subtração: a primeira se refere à diferença das médias da variável de resultado entre os períodos anterior e posterior ao programa, para o grupo de tratamento e para o de controle, e a segunda se refere à diferença da primeira diferença calculada entre esses dois grupos.
- A operacionalização do método requer a existência de informações para ambos os grupos para pelo menos um período de tempo antes e um período depois do programa.

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

a) Diferenças em Diferenças (DID)

- Embora essa demanda por dados torne a aplicação do procedimento mais custosa, o método possui uma série de vantagens relativa a outros métodos não experimentais.
- Em particular, uma das maiores vantagens é que ele é capaz de lidar com o viés de seleção associado a certo tipo de características não observáveis dos indivíduos, especificamente aquelas que são invariantes no tempo.

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

a) Método do Escore de Propensão (MEP)

- O método denominado pareamento busca tornar o grupo tratado semelhante a um grupo de não tratados, com base em um conjunto amplo de características observadas dos indivíduos.
- **Existem diversas maneiras de tornar os grupos de tratamento e controle semelhantes em características observadas, sendo a mais utilizada conhecida como método do escore de propensão (MEP).**
- **Nesse método, assume-se que a probabilidade de um indivíduo participar do programa depende apenas de um conjunto de características observadas, que estão disponíveis nos dados a serem utilizados pelo avaliador.**

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

a) Método do Escore de Propensão (MEP)

- Estimada essa probabilidade, o pareamento dos grupos é realizado em termos das semelhanças dessa probabilidade estimada.
- A ideia por trás do MPE é que, quando a probabilidade de participação é similar entre os indivíduos tratados e não tratados, os dois grupos tornam-se comparáveis, o que permite identificar o efeito causal da intervenção.
- A hipótese de que apenas as características observáveis são suficientes para contornar o viés de autosseleção não é passível de ser testada diretamente nos dados.

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

a) Regressão Descontínua (DRD)

- Muitos programas e políticas definem sua população-alvo com base em critérios que estabelecem valores de corte para quem pode ser ou não elegível a receber ou participar das ações da intervenção.
- Por exemplo, o programa Bolsa-Família estabelece que somente a família cuja renda per capita está abaixo de certa linha de corte pode receber a transferência monetária prevista pelo programa.
- Outro exemplo é o programa de Jovens Aprendizizes do governo federal, que fixa uma idade abaixo da qual um jovem estudante pode ser contratado com benefício fiscal por uma empresa no Brasil.

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

a) Método do Escore de Propensão (MEP)

- O método denominado regressão descontínua (ou design de regressão descontínua) procura explorar a existência desse tipo de descontinuidade nos critérios de elegibilidade dos programas para criar os grupos de tratamento e controle.
- A ideia básica é que os indivíduos não elegíveis que estão próximos à linha de corte (por exemplo, acima dela) são comparáveis em características observáveis e não observáveis aos indivíduos elegíveis que também estão próximos a essa linha (por exemplo, abaixo dela).

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

a) Método do Escore de Propensão (MEP)

- Como a única diferença entre esses dois grupos de cada lado do valor de corte é o fato de um poder participar do programa e outro não, o problema de autosseleção fica, pelo menos em princípio, resolvido.
- Assim, sob a hipótese de que não há uma descontinuidade marcante na variável de interesse em torno da linha de corte para o grupo de não elegíveis, a diferença nessa variável entre os dois grupos fornece o impacto do programa.

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

a) Método do Escore de Propensão (MEP)

- Embora seja um método que só pode ser empregado para avaliar o impacto de intervenções que possuam linhas de corte, ele é considerado bastante consistente para estimar o efeito de programas nessa categoria, pelo menos para os indivíduos que se situam em torno da linha de corte estabelecida pelas regras dos programas.

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

a) Método das Variáveis Instrumentais (VI)

- O método de variáveis instrumentais procura encontrar alguma variável (instrumento) que influencie a decisão de participar do programa e ao mesmo tempo não seja correlacionada com as características não observadas dos indivíduos que afetam a variável de resultado.
- Como é capaz de afetar a participação no programa, mas não está correlacionada com as características não observadas, essa variável corrige o problema de autosseleção, permitindo assim identificar o efeito causal de interesse.

Como contornar o viés de autosseleção (o arcabouço de resultados potenciais)

a) Método das Variáveis Instrumentais (VI)

- Naturalmente, a existência de uma variável com essas características vai depender do contexto no qual o programa está inserido e da disponibilidade dos dados.
- Ademais, tipicamente, não é possível testar a validade da hipótese de que o instrumento não é de fato correlacionado com as características não observáveis dos indivíduos.
- De todo modo, o método de variáveis instrumentais é uma abordagem bastante geral dentro da área de econometria, sendo também bastante empregada na área de avaliação de impacto.

Método de Aleatorização

Introdução

- O principal problema da área de avaliação é encontrar um grupo de controle que represente adequadamente o contrafactual do grupo tratado, ou seja, um grupo de comparação que seja o mais próximo possível do que teria ocorrido com o grupo tratado, caso esse não tivesse recebido a intervenção.
- Embora a situação contrafactual não tenha como ser concretamente observada, ela pode ser bastante bem aproximada pelo uso do método de aleatorização.
- Esse método, muitas vezes chamado de método experimental, é baseado na seleção aleatória dos indivíduos, famílias ou até mesmo comunidades inteiras que farão parte dos grupos de tratamento e controle.

Introdução

- Quando bem implementado, o mecanismo da aleatorização fornece o balanceamento necessário tanto das características observadas quanto das não observadas das unidades que compõem os dois grupos.
- Portanto, por construção, o método permite criar uma situação na qual não há correlação entre ser ou não tratado e os atributos das unidades de observação. O problema do viés de autosseleção fica assim resolvido.
- A avaliação aleatorizada é utilizada em diversos campos, sendo considerado o procedimento de referência para se estabelecer causalidade e medir o impacto de vários tipos de tratamento.

Introdução

- Por exemplo, ele é amplamente empregado nas áreas de medicina e farmacêutica, em que medicações e procedimentos médicos são avaliados com base na aleatorização dos indivíduos que receberão ou não o tratamento em avaliação.
- Na realidade, tamanha é a importância do método de aleatorização, que o seu uso é obrigatório para se comprovar a efetividade (e a existência de eventuais efeitos adversos) dos medicamentos que os laboratórios farmacêuticos pretendem disponibilizar para uso humano.

Introdução

- Além de ser o “padrão-ouro” para determinar a causalidade de um tratamento, o método de aleatorização tem outro apelo importante: sua simplicidade de cálculo.
- De fato, para obtermos o impacto da intervenção com base nesse método, basta subtrairmos a média da variável de interesse do grupo de tratamento da média correspondente calculada para o grupo de controle.
- Em condições ideais, essa simples diferença de médias fornece uma estimativa do efeito causal do tratamento sobre os tratados. Esse atributo dos experimentos sociais é um atrativo importante, pois torna o método mais transparente para os gestores do programa e para o público em geral.

Introdução

- Contudo, apesar de suas diversas vantagens, o método de aleatorização não está isento de problemas. Como veremos, a maior parte deles surge de situações práticas que ocorrem após a aleatorização dos indivíduos entre os grupos experimentais.
- **Os principais exemplos são o não comparecimento às atividades do programa por parte de uma parcela do grupo selecionado para o tratamento e a substituição da intervenção por outros programas similares por parte de uma fração do grupo de controle.**

Introdução

- Em ambos os casos, pode haver uma mudança de composição das características entre grupos experimentais, que foram selecionados a priori de forma aleatória, porém terminaram influenciados a posteriori por decisões de participação ou não no programa em avaliação ou em similares.
- Problemas desse tipo podem causar vieses não negligenciáveis na estimação do efeito que se pretendia isolar com o experimento aleatório.

Validade Interna e Externa

- A validade interna questiona se o desenho da avaliação identifica corretamente o efeito causal de um programa.
- Quando um estudo possui validade interna, seus resultados conseguem isolar o verdadeiro impacto do programa, isto é, estão livres da influência de outros fatores que não o próprio efeito da intervenção.
- Quando bem realizadas, as avaliações baseadas em experimentos aleatórios criam um grupo de controle que é estatisticamente equivalente ao grupo de tratamento antes do início do programa.

Validade Interna e Externa

- Em situação ideal, qualquer diferença observada entre os grupos após a intervenção é atribuída ao efeito causal do programa.
- Tem-se então que a seleção aleatória dos grupos de tratamento e controle assegura a validade interna de uma avaliação. Ela garante que os resultados da avaliação sejam válidos dentro do contexto específico daquela avaliação.
- Já a validade externa é a capacidade de estender os resultados para fora do contexto da avaliação.

Validade Interna e Externa

- Em outras palavras, uma avaliação possui validade externa se for possível generalizar as estimativas do impacto do programa avaliado para outras populações e outros momentos no tempo.
- Uma forma de garantir a validade externa dos resultados em certo ponto no tempo é que a amostra utilizada na avaliação seja representativa da população-alvo do programa naquele momento.
- Normalmente, quando a amostra é obtida por métodos de amostragem aleatória da população-alvo, essa representatividade é assegurada.

Validade Interna e Externa

a) Validade Interna

- A característica básica de um experimento aleatório é o uso de um mecanismo que seleciona aleatoriamente os indivíduos a participar ou não do programa.
- Quando essa seleção aleatória é feita de forma correta, cada indivíduo pertencente ao conjunto de participantes potenciais tem a mesma probabilidade de ser selecionado a participar da intervenção, independentemente de suas características pessoais.

Validade Interna e Externa

a) Validade Interna

- Em outras palavras, quando o processo de aleatorização é bem realizado, devemos esperar que os grupos de tratamento e controle sejam semelhantes tanto nas suas características observadas quanto nas não observadas.
- A única diferença que se espera entre os dois grupos é que o grupo de tratamento participe do programa, ao passo que o de controle não passe pela intervenção.

Validade Interna e Externa

a) Validade Interna

- A ideia de que deve haver um balanceamento de características entre os dois grupos aleatorizados é bastante natural.
- Para ver isso, considere uma situação hipotética em que há 2.000 indivíduos elegíveis para frequentar um programa de treinamento básico para o uso de planilhas eletrônicas em computadores.
- Suponhamos que, nesse total de duas mil pessoas, 1.200 (60%) sejam homens, 1.000 (50%) tenham cabelo preto e 400 (20%) sejam hábeis com números. Admita que esta última característica seja difícil de medir e, nesse sentido, torne-se não observada para nós.

Validade Interna e Externa

a) Validade Interna

- O que deveríamos esperar do balanceamento dessas características, se os 2.000 potenciais participantes fossem aleatorizados em dois grupos com 1.000 indivíduos em cada um?
- É intuitivo esperar que cada grupo fosse composto de aproximadamente 60% de homens, 50% de pessoas de cabelo preto e 20% com habilidades com números, ou seja, devemos esperar que os dois grupos sejam semelhantes entre si (e com o conjunto total de indivíduos).

Validade Interna e Externa

a) Validade Interna

- Note que essa semelhança deve se dar tanto nas características observadas (sexo e cor do cabelo) quanto nas não observadas (habilidade com números, no nosso exemplo).
- Isso é importante, pois é a aleatorização que garante que, em média, os dois grupos sejam similares em todas as dimensões.
- Na realidade, é essa equivalência entre os grupos que permite utilizar o grupo de controle como contrafactual para o grupo de tratamento. Esse é o principal motivo que faz com que o método de aleatorização seja considerado o padrão-ouro de estudos de efeitos causais.

Validade Interna e Externa

a) Validade Interna

- Em muitas áreas, inclusive, a validade interna de uma avaliação de impacto só é aceita se esta for baseada na aleatorização dos que vão ou não receber o tratamento que se quer estudar.
- Por exemplo, na área de farmacêutica, os medicamentos só são autorizados para uso geral se os laboratórios conduzirem com sucesso um experimento em que o grupo que recebe a medicação e o que não recebe (ou recebe um placebo) são selecionados de forma aleatória.

Validade Interna e Externa

a) Validade Externa

- Sem a validade interna de uma avaliação, pouca utilidade ela terá para revelar o verdadeiro efeito da intervenção.
- No entanto, mesmo quando a validade interna está assegurada, pode ocorrer de a avaliação ter sido realizada para um grupo de pessoas que não é representativo da população-alvo do programa.
- Por exemplo, se a avaliação de um programa nacional foi feita somente para uma localidade (bairro, cidade ou estado), não é possível generalizar os resultados da avaliação para o país como um todo.

Validade Interna e Externa

a) Validade Externa

- Em casos como esses, não se pode garantir a validade externa da avaliação, ou seja, os resultados não são extensíveis a outros grupos ou contextos.

Validade Interna e Externa

a) Validade Externa

- A validade externa é assegurada quando a amostra de indivíduos que é utilizada na avaliação é representativa da população de elegíveis a receber o programa.
- Por exemplo, a avaliação de impacto de um programa de qualificação profissional para jovens de ambos os sexos terá validade externa se a avaliação for baseada numa amostra representativa desses grupos demográficos.
- Naturalmente, se a amostra contiver indivíduos adultos ou for formada somente por homens, as conclusões da avaliação não podem ser generalizadas para a toda a população-alvo do programa.

Validade Interna e Externa

a) Validade Externa

- No entanto, não basta que a amostra só contenha indivíduos da população-alvo; é necessário também que os indivíduos da amostra representem acuradamente essa população.

Validade Interna e Externa

a) Validade Externa

- Na prática, a forma mais utilizada para garantir a representatividade de uma amostra é que ela seja uma amostra aleatória da população que ela pretende representar.
- As mais importantes pesquisas socioeconômicas conduzidas pelo IBGE e outros institutos oficiais de estatística do mundo são baseadas em amostras aleatórias retiradas da população dos países.
- No caso de uma amostra para fins de avaliação de impacto de um programa, a representatividade da população de elegíveis está assegurada se a amostra a ser empregada na avaliação for escolhida de forma aleatória a partir dessa população.

Validade Interna e Externa

a) Validade Externa

- Note que a aleatorização que garante a validade externa de uma avaliação não assegura a validade interna dessa avaliação, e vice-versa.
- Se a avaliação utiliza uma amostra aleatória da população-alvo do programa, mas a participação no tratamento não é realizada por aleatorização, então a amostra deve ser representativa da população de elegíveis, porém os resultados da avaliação podem não ter validade interna.

Validade Interna e Externa

a) Validade Externa

- Distintamente, caso a amostra não seja aleatória da população- alvo, mas a participação no programa seja aleatorizada entre os grupos de tratamento e controle, os impactos computados pela avaliação serão internamente válidos, porém não terão validade externa.

Custos da Aleatorização

- Durante as últimas décadas, o método experimental vem sendo cada vez mais utilizado em avaliações de impacto de programas sociais.
- Esse crescimento tem sido observado tanto em países ricos quanto pobres, e tem coberto intervenções nas áreas de educação, saúde, treinamento profissional, pobreza, entre outras.
- Contudo, apesar desse crescimento, o método de aleatorização (ainda) está longe de ser predominantemente utilizado em avaliações de impacto de programas sociais.

Custos da Aleatorização

- Mas por que ele não é o mais utilizado, visto que é o único que garante com poucas hipóteses a validade interna das avaliações de impacto? A resposta a essa pergunta reside em três conjuntos de fatores: éticos, financeiros e técnicos.

Custos da Aleatorização

- Muitos argumentam que não é ético negar a participação em certo programa social com base num mecanismo de aleatorização.
- Afinal, dizem os que se opõem ao método, quando se trata de programa social não é eticamente correto deixar pessoas fora do programa com base num simples “lançamento de moeda”.
- Sem dúvida, não se cogita utilizar aleatorização para casos de intervenções de ampla escala (por exemplo, programas de vacinação), ou para públicos-alvo muito específicos (por exemplo, pessoas com algum tipo de deficiência).

Custos da Aleatorização

- No entanto, há pelo menos duas situações importantes em que esse problema ético não se coloca.
- A primeira, mais comum, ocorre quando existe escassez de vagas para atender toda a demanda ou público-alvo do programa. Nessa situação, pode-se inclusive justificar o uso da aleatorização como o critério mais justo de alocação das vagas existentes entre os candidatos ao programa.
- De fato, se pensarmos bem, em vez de utilizar o tradicional critério de atendimento por ordem de chegada, em muitas situações pode ser mais correto oferecer as vagas disponíveis, garantindo a todos os elegíveis a mesma probabilidade de acesso a elas.

Custos da Aleatorização

- Se, por exemplo, há escassez relativa de vagas e não se conhece bem onde os elegíveis moram, parece mais justo receber todas as inscrições e selecionar os participantes via processo de aleatorização – por exemplo, retirando aleatoriamente os nomes de uma urna na presença dos inscritos – do que oferecer as vagas disponíveis para os primeiros que chegarem ao local de inscrição.
- A aleatorização das vagas, nesse sentido, não é recomendada para assegurar somente a validade interna da avaliação, mas também para garantir a igualdade de oportunidade de acesso aos serviços do programa.

Custos da Aleatorização

- A segunda situação na qual é mais fácil eticamente utilizar o método de aleatorização ocorre quando os recursos do programa (financeiros, de pessoal ou operacionais) são limitados e, apesar de se planejar atender toda a população-alvo ao final da implementação, só é possível incluir os elegíveis ao longo do tempo.
- Em casos como esse, ninguém será privado de participar do programa; o que se aleatoriza é somente a ordem com que os elegíveis entram no programa.

Custos da Aleatorização

- Ao longo do processo gradual de entrada no programa, os que ainda não tiverem participado servem como grupo de controle para os que já foram selecionados.
- Muitas vezes, essa aleatorização gradual não é realizada em nível individual ou familiar e sim em nível mais agregado, como bairros, comunidades ou municípios.

Custos da Aleatorização

- Exemplo bastante conhecido de aleatorização gradual foi o programa mexicano de transferência condicional de renda, Progresa (hoje em dia chamado Oportunidades), que teve a primeira parcela dos municípios considerados elegíveis selecionados de forma aleatória pelos gestores do programa.
- Mas não é só por razões éticas que os experimentos sociais não são realizados com mais frequência. Outro motivo importante é que eles podem ser dispendiosos em relação ao orçamento global do programa.
- Tipicamente, a avaliação aleatorizada de um programa social requer a mobilização de vários recursos, tais como o tempo dos gestores do programa, a contratação de pessoal técnico e a realização de pesquisas de campo com as unidades tratadas e de controle.

Custos da Aleatorização

- Este último componente requer o uso de questionários especialmente desenhados para a avaliação do programa em questão, que são geralmente aplicados em dois momentos do tempo: o primeiro tipicamente no momento da inscrição do programa, e o segundo alguns meses após o final da intervenção.
- Normalmente, é esse componente que requer o maior volume de recursos financeiros e, portanto, é o que merece mais atenção no planejamento da avaliação.
- Assim, apesar de todas as vantagens do método experimental, é importante contrapor os benefícios e custos de sua aplicação vis a vis a de outros métodos.

Custos da Aleatorização

- Há também razões de natureza técnica que dificultam o uso de experimentos aleatórios.
- Entre as mais importantes estão: o não comparecimento ao programa de uma parcela dos indivíduos do grupo de tratamento; a possibilidade de que pelo menos parte do grupo de controle acabe encontrando uma maneira de entrar no programa ou substituí-lo por um semelhante; e a perda de observações por atrito amostral, isto é, a perda das informações sobre unidades tratadas e de controle ao longo do tempo.

Dificuldades Técnicas da Aleatorização

- Apesar de os experimentos sociais apresentarem importantes vantagens sobre outros métodos de avaliação de impacto, eles não estão livres de problemas.
- **Quatro são as principais dificuldades técnicas que aparecem na realização dos experimentos aleatorizados.**

Dificuldades Técnicas da Aleatorização

- A primeira é o não comparecimento de uma parcela dos indivíduos selecionados para o grupo de tratamento para realizar o programa.
- A segunda, muito frequente, é a substituição do programa em avaliação por outros semelhantes por parte dos indivíduos do grupo de controle.
- A terceira dificuldade, chamada atrito amostral, tipicamente aparece quando há coleta de dados em mais de um ponto do tempo e ocorre perda de observações de um conjunto de indivíduos da amostra entre períodos.
- A quarta dificuldade surge quando o próprio programa gera externalidades para os não tratados – incluindo os membros do grupo de controle –, o que pode distorcer os efeitos computados.

Dificuldades Técnicas da Aleatorização

a) Não Comparecimento

- **Um dos problemas mais comuns de um experimento social aleatório ocorre quando, após terem sido selecionados para participar do programa, alguns indivíduos do grupo de tratamento não aparecem de fato para realizar as atividades previstas na intervenção.**
- **O não comparecimento desse subgrupo altera o grupo de tratamento, o que pode comprometer a identificação dos impactos do programa, especialmente se a desistência estiver associada com características não observadas dos desistentes.**

Dificuldades Técnicas da Aleatorização

a) Não Comparecimento

- Por exemplo, se os menos motivados entre os selecionados para o programa são os que não comparecem à composição do grupo de tratamento deixa de ser parecida com a do grupo de controle nessa característica.
- Assim, se motivação for um determinante importante da variável de resultado (por exemplo, busca por emprego), a diferença de composição que aparece a posteriori entre os dois grupos irá comprometer o isolamento do que é efetivamente o impacto do programa.

Dificuldades Técnicas da Aleatorização

a) Não Comparecimento

- Uma forma de olhar para esse problema é que ocorre uma distinção entre a oferta e a participação propriamente dita no programa.
- Essa distinção, que ocorre com vários programas sociais e não somente com experimentos aleatorizados, pode criar diferença entre o impacto que os gestores intencionavam que o programa tivesse para o público elegível em geral e o impacto que de fato ocorreu para o público efetivamente tratado.
- Na realidade, essa diferença de impactos tem sido amplamente discutida na literatura de avaliação de impacto e pode ser entendida dentro do arcabouço dos experimentos aleatórios.

Dificuldades Técnicas da Aleatorização

a) Substituição

- O problema inverso também costuma ocorrer com os indivíduos que compõem o grupo de controle.
- **Idealmente, num experimento social, nenhum membro desse grupo deveria receber os serviços do programa ou de substituto próximo dele.**
- **Contudo, na prática, não é incomum que uma parcela do grupo acabe entrando no programa em avaliação ou em algum programa similar.**
- Quanto mais substituível a intervenção sendo avaliada, maior a possibilidade de que os controles busquem tratamento parecido em algum outro lugar.

Dificuldades Técnicas da Aleatorização

a) Substituição

- A principal consequência dessa substituição do programa é que o grupo de controle experimental deixa de corresponder ao contrafactual do estado de não tratamento.
- Como decorrência, a média da variável de resultado para o grupo de controle deixa de fornecer a média que se buscava com o experimento aleatório.
- Com isso, a simples diferença das médias calculadas para os grupos de tratamento e controle experimentais não isola mais o impacto do programa.

Dificuldades Técnicas da Aleatorização

a) Atrito

- Em quase todas as pesquisas que coletam dados de campo em mais de um momento do tempo, há perda de informações sobre alguns indivíduos que não são encontrados nos períodos subsequentes ao primeiro.
- Há diversos motivos que levam ao aparecimento desse atrito amostral, mas muitos deles não são aleatórios, ou seja, podem estar relacionados às características (não observadas) dos indivíduos ou ao impacto de interesse.
- Normalmente, quando o atrito não é aleatório, a estimação do efeito causal do programa por meio da comparação entre tratados e controles torna-se potencialmente enviesada.

Dificuldades Técnicas da Aleatorização

a) Atrito

- A aparição desse tipo de viés é mais provável quando a taxa de atrito difere entre o grupo de tratamento e o grupo de controle.
- Na prática, essa é a situação mais comum, com a taxa do grupo de controle, sendo tipicamente mais elevada que a do grupo de tratamento.
- Muitas vezes, a negação do programa aos controles gera uma reação de recusa por parte deles em fornecer informações aos entrevistadores do programa.
- Em alguns casos, os controles, quando se veem na situação de não encontrar um programa alternativo perto de onde moram, mudam-se para outras localidades, tornando mais difícil a coleta subsequente de informações.

Dificuldades Técnicas da Aleatorização

a) Atrito

- Como já mencionado, é sempre boa prática comparar as características observadas dos grupos de tratamento e controle, antes e após o tratamento.
- Essa comparação pode ser particularmente informativa quando a taxa de atrito é substancial, pois, se aparecerem diferenças marcantes entre os dois grupos nas características observadas, é provável que o atrito tenha alterado também a composição dos atributos não observados entre os grupos.
- Tipicamente, a resolução do problema do viés de atrito amostral requer o uso complementar de hipóteses comuns aos métodos não experimentais.

Dificuldades Técnicas da Aleatorização

a) Externalidades

- Os experimentos aleatórios também podem estar associados a dois outros tipos de problemas.
- **O primeiro é um efeito de transbordamento do programa sobre o grupo de controle, ou seja, uma externalidade (positiva ou negativa) que é gerada pelo próprio tratamento sobre a variável de resultado do grupo de controle.**
- Por exemplo, se os grupos de tratamento e controle de um programa de combate à obesidade infantil são escolhidos aleatoriamente dentro das escolas, é provável que o tratamento acabe afetando indiretamente o grupo de controle, uma vez que as crianças tratadas podem influenciar o comportamento alimentar das crianças do grupo de controle (dentro ou fora da escola).

Dificuldades Técnicas da Aleatorização

a) Externalidades

- Em casos como esse, é mais recomendado realizar a aleatorização dos grupos experimentais entre escolas, preferencialmente guardando certa distância entre as selecionadas para cada grupo.
- Esse tipo de cuidado deve ser sempre tomado com intervenções de saúde, educacionais, alimentares, entre outras, em que podem ocorrer externalidades devido ao contato físico ou de comunicação dos indivíduos tratados com os não tratados.

Dificuldades Técnicas da Aleatorização

a) Externalidades

- **Finalmente, o último problema que precisa ser levado em consideração é a ocorrência de efeitos de equilíbrio geral, isto é, efeitos que o programa pode ter sobre o funcionamento da economia do local onde ele ocorre.**
- Esses efeitos são mais comuns quando a escala do programa é grande relativamente ao tamanho da economia.

Dificuldades Técnicas da Aleatorização

a) Externalidades

- Por exemplo, o programa Bolsa Família atende a uma proporção que pode chegar a quase 90% da população de vários municípios pequenos no Brasil.
- Em municipalidades como essas, as transferências de renda às famílias beneficiadas pelo programa pode gerar impacto significativo sobre a demanda por diversos bens e serviços dessas localidades, resultando em mudanças de preços e salários.

Dificuldades Técnicas da Aleatorização

a) Externalidades

- **Esse tipo de efeito agregado da intervenção tende a alterar o próprio estado do contrafactual de não tratamento, tipicamente concebido para uma situação na qual não há mudanças na economia decorrentes do próprio programa.**
- Os efeitos de equilíbrio geral podem provocar uma reação dos agentes sobre suas decisões futuras em relação à variável de interesse, pelo menos em longo prazo.
- Por exemplo, um amplo programa de concessão de bolsas de estudos universitárias pode ter efeito positivo sobre a entrada dos beneficiários na universidade. No entanto, devido a sua escala, o programa também tende a alterar o retorno à educação superior, o que pode afetar as decisões das pessoas de fazer curso superior no futuro.

Dificuldades Técnicas da Aleatorização

a) Externalidades

- A maior parte das avaliações de impacto não leva em conta esse tipo de efeitos indiretos do programa na economia e nas decisões dos agentes.
- Essa prática não é recomendada, especialmente nos casos de intervenções de larga escala.
- A forma típica de resolver esse problema é lançar mão de um modelo estrutural geral que incorpore de forma explícita o comportamento dos agentes.
- Após a estimação dos parâmetros do modelo – inclusive o referente ao impacto do programa –, é possível simulá-lo para conhecer os efeitos diretos e indiretos da intervenção.