

SINOPSE

O artigo faz uma apresentação introdutória da inferência causal e sua importância para a avaliação de políticas públicas, destacando sua importância para compreender se uma intervenção desencadeia efeitos que não ocorreriam na ausência dela. Em contraste com a inferência estatística, que busca estimar parâmetros e extrapolá-los para a população, a inferência causal visa determinar se o parâmetro estimado reflete a relação de causa e efeito que se deseja analisar. O artigo apresenta o conceito de contrafactual como uma ferramenta crucial para abordar questões causais e chama a atenção para a inadequação de comparações intuitivas do tipo antes *versus* depois e afetados *versus* não afetados no exame de questões de natureza causal. O texto destaca a importância não apenas de dados e procedimentos estatísticos, mas também de hipóteses nem sempre empiricamente testáveis. O objetivo final é promover a compreensão de inferências causais válidas e identificar abordagens malsucedidas.

Palavras-chave: inferência causal; avaliação de políticas públicas; contrafactual.

1 INTRODUÇÃO

Qual é o efeito de uma política pública? Essa pergunta, aparentemente simples, pode ter diferentes interpretações. Uma delas envolve a ideia de causalidade: a intervenção desencadeia um ou mais efeitos que poderiam não ocorrer na ausência dela. Responder a perguntas com essa interpretação – fundamentais quando se tem por objetivo produzir evidências úteis para informar políticas públicas – é o objetivo fundamental da inferência causal.

Note-se que se trata de um exercício de natureza bastante distinta, embora complementar, da inferência estatística. Nesta, o objetivo é obter uma estimativa de um parâmetro – uma correlação entre duas variáveis, por exemplo – e extrapolá-la para a população de interesse. No caso da inferência causal, o objetivo é determinar se o parâmetro estimado mede a relação de causa e efeito que se deseja analisar. Para isso, como será discutido ao longo deste artigo, faz-se necessário não apenas o uso de dados e procedimentos estatísticos, mas também de hipóteses, que nem sempre podem ser testadas empiricamente, sobre a questão causal que se deseja responder.

O propósito deste artigo é apresentar, de maneira bastante introdutória e intuitiva, o campo da inferência causal e a ideia de contrafactual, uma ferramenta conceitual poderosa e útil para pensar em questões de natureza causal. O objetivo aqui é provocar a leitora ou o leitor a pensar sobre a

1. DOI: <http://dx.doi.org/10.38116/bapi37art10>

2. Técnico de planejamento e pesquisa na Diretoria de Estudos e Políticas do Estado, das Instituições e da Democracia do Instituto de Pesquisa Econômica Aplicada (Diest/Ipea); e doutor em economia pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).
E-mail: flavio.carneiro@ipea.gov.br.

forma com que pensa sobre perguntas causais, para tornar-se capaz de compreender o que é uma inferência causal válida – e, não menos importante, identificar tentativas malsucedidas.³

2 PERGUNTAS CAUSAIS E A IDEIA DE CONTRAFCTUAL

Embora raciocinar sobre perguntas de natureza causal faça parte do dia a dia de qualquer ser humano, as heurísticas e comparações intuitivas que geralmente utilizamos para tentar responder a essas perguntas são frequentemente incapazes de levar a uma inferência causal válida. O erro mais comum é associar imediatamente correlação e causalidade, a ponto de a expressão “correlação não significa causalidade” ter se tornado lugar-comum. Não, o consumo de chocolate *per capita* não aumenta a probabilidade de um país ter cientistas premiados com o Nobel; tampouco o sol nasce por causa do canto do galo. É surpreendentemente elevada, contudo, a quantidade de vezes em que utilizamos intuitivamente comparações tão inválidas quanto esses exemplos caricatos para pensar sobre questões de natureza causal – levando, evidentemente, a respostas inválidas.

Um exemplo simples de pergunta causal nos permite identificar claramente essas comparações intuitivas e a razão pelas quais elas podem ser incapazes de levar a uma resposta adequada. Suponha que desejamos estimar o efeito causal de um aumento no número de policiais sobre a atividade criminal; dito de outra forma, queremos responder à seguinte pergunta: quanto mais policiais, menor a criminalidade? Ou ocorre o contrário?

Suponha ainda que temos uma base de dados para uma amostra de regiões – digamos, todos os municípios de um determinado Estado – e, para cada região, alguma medida de criminalidade em dois períodos (abril e maio), bem como o número de policiais em atividade em cada período. Além disso, suponha que em algumas regiões o número de policiais foi aumentado em maio, enquanto nas outras esse número se manteve constante nos dois meses.

A partir desses dados, podemos tentar estimar o efeito do aumento no efetivo policial sobre a criminalidade (uma correlação). Em geral, pensamos intuitivamente em duas possibilidades “ingênuas”. A primeira consiste em comparar a criminalidade média em abril e em maio nas regiões que tiveram seu efetivo policial aumentado – isto é, comparar os resultados observados antes e depois da intervenção. A segunda seria comparar a criminalidade média em maio das regiões que sofreram a intervenção com a das regiões cujo efetivo policial se manteve constante. Essas comparações são capazes de responder à pergunta causal?

Suponha que o resultado da primeira comparação (antes *versus* depois) indique uma relação inversa entre as grandezas comparadas: após o aumento do efetivo policial, as regiões passam a ter, em média, criminalidade menor. Contudo, isso não necessariamente significa que encontramos um efeito causal negativo (mais polícia causa menos crime), uma vez que a correlação observada pode ser decorrente do efeito de outras variáveis não incluídas na análise. Por exemplo, se a criminalidade for inversamente correlacionada com o nível de emprego e se este tivesse aumentado entre abril

3. Pelo escopo e tamanho limitados, este artigo não aborda uma série de pontos que serão tratados em versão ampliada (Carneiro, 2024). Por exemplo, a discussão sobre o conceito de causalidade, que aqui limita-se ao sentido contrafactual. Aqui se trata apenas da inferência causal como uma solução particular para o problema maior (e mais antigo) da causalidade; como colocado por Holland (1986), trata-se de identificar efeitos de causas, e não causas de efeitos. Além disso, não se discute a controvérsia, ainda ativa, sobre a importância de modelos estruturais para inferência causal, que remonta a Koopmans (1947). Por fim, não são abordados métodos para dados qualitativos – a esse respeito ver, por exemplo, King, Keohane e Verba (1994).

e maio, observaríamos uma queda na criminalidade mesmo se o aumento no efetivo policial não tivesse nenhum efeito causal.

Agora suponha que encontramos um efeito positivo na segunda comparação: regiões nas quais o número de policiais aumentou apresentaram maior criminalidade média em maio do que aquelas em que o efetivo permaneceu constante. Novamente, isso não necessariamente significa que encontramos um efeito causal (desta vez positivo: quanto mais polícia, mais crime). É possível que as autoridades tenham optado por reforçar o policiamento das regiões em que a criminalidade já era maior, ou que já estava em tendência de alta. Nesse caso, a causalidade atuaria no sentido contrário: quanto mais crime, mais polícia; a correlação observada, contudo, é a mesma, independentemente da direção do efeito causal.

Problemas como esses – a existência de viés de variável omitida, na primeira comparação ingênua, e de causalidade reversa, na segunda⁴ – fazem com que muitas das respostas intuitivas a questões de natureza causal sejam incapazes de levar a uma inferência causal válida. Não obstante, nos deparamos com a utilização de comparações desse tipo a todo momento; uma simples consulta ao noticiário revela a frequência de sua utilização para tentar – sem sucesso, obviamente – elucidar o efeito causal de políticas públicas, tanto ao comparar a situação antes e depois da implementação de determinada política, quanto ao comparar unidades afetadas por uma política com as não afetadas.

A razão pela qual essas comparações intuitivas são inadequadas para responder perguntas de natureza causal pode ser consubstanciada no conceito de contrafactual: o que teria acontecido com as unidades (regiões, em nosso exemplo) afetadas pela intervenção, caso a intervenção não tivesse acontecido, e tudo mais tivesse permanecido constante.

Nenhuma das duas comparações intuitivas é capaz de prover um bom contrafactual. Na primeira (antes *versus* depois), uma miríade de outros fatores pode estar atuando para confundir os efeitos da intervenção, de modo que nem tudo permanece constante, contradizendo a definição de contrafactual. Já na segunda (afetados *versus* não afetados), a comparação se dá entre unidades diferentes, o que também viola a definição de contrafactual.

O conceito de contrafactual, embora simples, pode ser bastante contraintuitivo na prática. O ponto-chave para uma inferência causal válida é a construção de um bom contrafactual; assim, na próxima seção, será apresentado, de maneira sucinta, um arcabouço conceitual bastante útil para sua compreensão, o modelo de resultados potenciais (Rubin, 1974).

3 O MODELO DE RESULTADOS POTENCIAIS

Inicialmente adotado a partir de Jerzy Neyman em 1923 para analisar experimentos (Neyman, 1990), mas popularizado depois de sua utilização para dados observacionais por Rubin (1974),

4. Ambos os problemas causam o que se costuma chamar de endogeneidade, diante do fato de que a variável de intervenção – neste caso, o aumento no número de policiais – é endógena, ou seja, influenciada por outra variável no modelo (o nível de criminalidade) – e não exógena, como se esperaria *a priori*.

o modelo de resultados potenciais é uma ferramenta simples e poderosa para pensar sobre questões causais e compreender o que é um contrafactual.⁵

A ideia central do modelo consiste em imaginar que cada unidade (que, a depender da aplicação, pode ser um indivíduo, uma firma, uma região etc.) possui duas⁶ ações possíveis, e cada uma dessas ações que a unidade pode tomar possui um resultado potencial associado. Por exemplo, um indivíduo pode tomar ou não um medicamento, e há um resultado potencial associado ao ato de tomar o medicamento, e outro ao ato de não o fazer.

Entre as ações possíveis, cada unidade só pode adotar uma; assim, apenas o resultado potencial associado à ação que de fato foi tomada será observado. O resultado potencial associado à ação que não foi realizada – que, portanto, não pode ser observado – é o contrafactual. Por exemplo, digamos que o indivíduo tomou o medicamento; o resultado potencial associado a essa ação pode ser observado examinando-se seu estado de saúde após tomar o medicamento, enquanto o resultado potencial associado à outra ação, isto é, o contrafactual – seu estado de saúde após não tomar o medicamento – nunca ocorre na realidade e, portanto, não pode ser observado.

Nesse arcabouço, o efeito causal de uma ação é definido como a diferença entre o resultado potencial associado a essa ação e o resultado potencial associado à outra ação; isto é, a diferença entre o resultado observado e o contrafactual. O leitor atento certamente já identificou o problema: o efeito causal depende diretamente de um resultado que não existe na realidade, e que, portanto, não pode ser observado.

Podemos utilizar esse arcabouço para analisar o exemplo introduzido na seção anterior, e identificar mais claramente os problemas associados às duas comparações ingênuas que, como vimos, não identificam corretamente o efeito causal da intervenção. Para manter a análise o mais simples possível, suponha que existam duas regiões, A e B, e que cada uma pode tomar duas ações possíveis – adotar um efetivo policial pequeno ou grande – em cada um dos dois períodos, a que denominaremos antes e depois. Suponha, ainda, que, associada a cada ação possível de cada região em cada período, há um resultado potencial: alguma medida de criminalidade que pode ser registrada. Finalmente, considere que no período “antes” ambas as regiões adotaram efetivo pequeno, e que no período “depois” a região A optou por mudar sua ação, passando a ter um efetivo grande, enquanto a região B manteve o efetivo pequeno.

Podemos examinar a intervenção ocorrida em A – o aumento no número de policiais – para tentar responder, nesse cenário hipotético, nossa questão causal: quanto mais policiais, mais crime? Ou o contrário é verdade? Para dar substância ao exemplo, podemos definir quais são os resultados observados e potenciais, identificar qual é o contrafactual da intervenção, e calcular o efeito causal.⁷

5. Evidentemente, o modelo de resultados potenciais não é o único arcabouço utilizado para pensar sobre questões causais; sua popularidade reside, em grande medida, em sua simplicidade (se comparado, por exemplo, com modelos estruturais de equações simultâneas) e aplicabilidade à inferência causal em diferentes contextos. Outro arcabouço que tem ganhado popularidade recentemente, e não é coberto neste artigo por questões de espaço, se baseia na utilização de gráficos diretos acíclicos para traçar relações causais; uma apresentação simples desse arcabouço se encontra em Pearl e Mackenzie (2018), e uma comparação das duas abordagens pode ser encontrada em Imbens (2020).

6. O modelo pode ser imediatamente estendido para considerar mais ações. Optou-se, aqui, pela versão binária, por motivos didáticos e para manter a simplicidade na exposição.

7. A expressão “efeito causal” é amplamente adotada na literatura e, embora possa parecer redundante, é útil para destacar a oposição ao coeficiente de uma variável em uma regressão, comumente chamado de efeito, mas que não necessariamente possui interpretação causal.

TABELA 1
Resultados observados

Efetivo	Período: antes		Efetivo	Período: depois	
	Região			Região	
	A	B		A	B
Pequeno	34	24	Pequeno	-	20
Grande	-	-	Grande	22	-

Elaboração do autor.

A tabela 1 apresenta os resultados observados de ambas as regiões nos dois períodos. Isso é o que estaria disponível na base de dados de uma pesquisadora que deseja analisar uma situação como essa: no período “antes”, os resultados potenciais associados ao efetivo pequeno para ambas as regiões (34 para A e 24 para B), e, no período “depois”, os resultados associados ao efetivo grande para a região A (22) e ao efetivo pequeno para a região B (20). Ela não terá acesso aos demais resultados potenciais, associados às ações que não foram tomadas; crucialmente, ela não poderá observar o contrafactual da intervenção: o resultado associado ao efetivo pequeno no período “depois” para a região, fundamental para calcular o efeito causal.

Agora suponha que tenhamos acesso a todos os resultados potenciais associados a cada tríade ação-unidade-período, apresentados na tabela 2. Pode se notar, imediatamente, que a região A tem maior criminalidade em todas as situações possíveis: seus resultados potenciais são sempre maiores que os análogos na região B. Além disso, é possível perceber ainda que, por alguma razão, o período “depois” experimentou uma redução da criminalidade *vis-à-vis* o período “antes”.

TABELA 2
Resultados potenciais

Efetivo	Período: antes		Efetivo	Período: depois	
	Região			Região	
	A	B		A	B
Pequeno	34	24	Pequeno	30	20
Grande	26	18	Grande	22	14

Elaboração do autor.

De posse de todos os resultados potenciais, o leitor pode facilmente identificar o contrafactual da intervenção (realçado na tabela), e calcular seu efeito causal – que, como visto, é dado pela diferença entre o resultado observado e o contrafactual: $22 - 30 = -8$. O aumento do efetivo policial de pequeno para grande causa, na região A, uma redução de oito pontos na medida de criminalidade.

Ressalte-se, contudo, que isso só é possível porque, nesse exemplo artificialmente construído, temos acesso a todos os resultados potenciais. Na realidade, obviamente, isso não acontece, e a pesquisadora não é capaz de calcular diretamente o efeito causal. O que ela obteria caso adotasse alguma das comparações ingênuas?

A primeira consiste em comparar o resultado observado para a região A nos dois períodos: $22 - 34 = -12$; a segunda, comparar o resultado das duas regiões no período depois: $22 - 20 = 2$. Em ambos os casos, a distância entre a estimativa e o efeito causal (o viés) é considerável: a primeira

superestima a intensidade do efeito (encontrando uma redução maior do que a verdadeira), enquanto a segunda o efeito a ponto de inverter seu sinal.

Podemos ainda fazer uma decomposição simples para identificar a origem de cada viés. O resultado da comparação antes *versus* depois é dado por:

$$-12 = 22 - 34 = 22 - 30 + 30 - 34 = (-8) + (-4) = (\text{efeito causal}) + (\text{viés de variável omitida})$$

Ou seja, o efeito total estimado é a soma do efeito causal com um outro termo que nada tem a ver com este: a diferença entre os resultados potenciais que seriam observados para a região A nos dois períodos, caso a intervenção não tivesse ocorrido – que denominamos “viés de variável omitida”.

A comparação afetados *versus* não afetados resulta em:

$$2 = 22 - 20 = 22 - 30 + 30 - 20 = (-8) + (10) = (\text{efeito causal}) + (\text{viés de seleção})$$

Novamente, o efeito estimado acresce ao efeito causal um termo adicional: a diferença entre os resultados potenciais que seriam observados no segundo período para as duas regiões, na ausência da intervenção – que denominamos “viés de seleção”.

Assim, para calcular o efeito causal de uma intervenção sobre uma unidade seria necessário comparar diretamente o resultado ocorrido com o contrafactual – uma tarefa impossível, uma vez que este último é, por definição, não observável (pois não ocorreu na realidade).

Extensa literatura tem sido construída com o objetivo de lidar com esse problema fundamental da inferência causal: já que não é possível calcular o efeito causal, podemos tentar estimá-lo, tentando reduzir ao máximo essas duas fontes de viés. Um conjunto de técnicas tem sido desenvolvido para lidar com o desafio de construir um contrafactual adequado, que emule da melhor maneira possível o contrafactual verdadeiro – seja buscando unidades não afetadas que sejam similares às unidades afetadas para comparação, ou enfocando intervenções exógenas aos resultados de interesse. A próxima seção apresenta, de maneira sucinta, algumas dessas técnicas e as hipóteses subjacentes a elas.

4 TÉCNICAS DE INFERÊNCIA CAUSAL

4.1 O poder da aleatorização

O método considerado ideal para estimar o efeito causal é a utilização dos chamados experimentos aleatorizados controlados (*randomized controlled trials* – RCTs), análogos aos utilizados no desenvolvimento de medicamentos, por exemplo.

Suponha que existe um grande número de regiões, diferentes em várias dimensões que podem causar viés (características demográficas, econômicas etc.). Se dividirmos essas regiões de maneira aleatória entre dois grupos, A e B, a média de qualquer variável será igual entre os dois grupos.

Para ilustrar o poder da aleatorização, considere a tabela 3, que mostra a média de homicídios em 2018 de municípios dos estados de Rio de Janeiro e São Paulo. No painel superior, esses municípios estão divididos por estado – uma divisão claramente não aleatória. Nota-se uma diferença clara e estatisticamente significativa (*p-valor* 0.000) entre as médias: municípios do Rio de Janeiro tiveram, em média, um número muito maior de homicídios que os de São Paulo.

TABELA 3
Média de homicídios por município (2018)

Divisão não aleatória (por Unidade da Federação)				
	Rio de Janeiro	São Paulo	Diferença	Pr(T > t)
Média de homicídios	72,10	9,37	62,72	0,000
Divisão aleatória				
	Grupo A	Grupo B	Diferença	Pr(T > t)
Média de homicídios	23,10	17,40	5,70	0,532

Fonte: Datasus.

Já no painel inferior, os mesmos municípios foram divididos aleatoriamente em dois grupos, independentemente do estado. Os valores por município são exatamente os mesmos, mas as médias dos dois grupos são bastante próximas, e estatisticamente indistinguíveis (*p-valor* 0.53).

A grande vantagem desse procedimento para a inferência causal é que ele vale para qualquer variável, incluindo os resultados potenciais, que serão iguais, em média, entre os dois grupos. Assim, alterando o exemplo da seção anterior, obteríamos algo como o ilustrado na tabela 4.

TABELA 4
Resultados potenciais médios

Efetivo	Período: antes		Período: depois	
	Região		Região	
	A	B	A	B
Pequeno	29	29	25	25
Grande	22	22	18	18

Elaboração do autor.

O resultado é que o grupo de controle B passa a ser um excelente contrafactual para o que teria ocorrido com o grupo de tratamento A na ausência da intervenção: o viés de variável omitida é igual entre os dois grupos ($25 - 29 = -4$), e o viés de seleção desaparece ($25 - 25 = 0$). Podemos estimar o efeito causal médio comparando os resultados médios entre os dois grupos no período depois da intervenção: $18 - 25 = -7$.

Note que isso decorre não do fato de realizarmos algum experimento controlado em laboratório que fosse capaz de “controlar” ou “isolar” o efeito de outras variáveis, mas sim do poder da aleatorização das unidades entre os dois grupos (tratamento e controle), que iguala, na média, quaisquer características entre ambos – e, em particular, iguala os resultados potenciais médios dos dois grupos, de modo que o resultado observado para o grupo de controle pode ser utilizado no lugar do contrafactual (o resultado potencial não observado do grupo de tratamento).

Embora seja a regra em ciências naturais e medicina, a utilização de RCTs em ciências sociais não é tão simples; a intervenção controlada pode ser impossível, logisticamente inviável, ou antiética. Ainda assim, seu uso não é novo e vem crescendo em diversas áreas, tais como educação, políticas de transferência de renda e combate à pobreza, e segurança pública, por exemplo – um panorama recente dessa literatura pode ser encontrado em Mize e Manago (2022).

Além do alto custo e das questões logísticas e éticas, os RCTs envolvem uma série de desafios técnicos que podem dificultar ou mesmo inviabilizar sua implementação. Por exemplo, a falta de estabilidade nos grupos de tratamento e controle (quando algum participante abandona o estudo, por exemplo) ou a presença de externalidades (isto é, quando a intervenção sobre uma unidade pode afetar outras unidades) podem tornar a aleatorização inválida, levando a grupos de tratamento e controle com resultados potenciais médios distintos, e reintroduzindo as duas fontes de vies.

Ademais, embora os RCTs possuam validade interna – isto é, sejam capazes de identificar corretamente o efeito causal, desde que a aleatorização seja bem-feita –, podemos questionar sua validade externa – o quanto podemos generalizar as conclusões obtidas por um experimento realizado em um contexto específico para outras situações, ou para uma escala de implementação distinta. Dito de outra forma, mesmo que um RCT nos dê uma estimativa correta do efeito causal de uma intervenção, nada garante que este será o mesmo em outra região, ou sobre um grupo sociodemográfico distinto, por exemplo, e se os resultados de um projeto piloto continuarão válidos em uma implementação ampla da intervenção.

4.2 Inferência causal com dados observacionais – uma breve introdução

Uma limitação ainda mais evidente dos RCTs advém do fato de que eles exigem algum grau de manipulação da intervenção por parte do pesquisador. Isso restringe sobremaneira o escopo das perguntas de pesquisa que podem ser respondidas: frequentemente, estamos interessados em avaliar uma política sobre a qual não temos controle, seja porque é da alçada de outra entidade ou mesmo porque já foi implementada. Torna-se necessário, portanto, dispor de técnicas de inferência causal aplicáveis a dados observacionais – isto é, que não foram gerados a partir de experimentos controlados (tais como pesquisas populacionais ou amostrais, ou dados administrativos).

O desafio, como apontado, é obter um bom contrafactual, maximizando a similaridade entre “tratados” e “controles”, ou buscando uma intervenção que seja plausivelmente exógena ao resultado que se quer analisar. A literatura sobre inferência causal tem desenvolvido uma série de técnicas que combinam, em alguma medida, esses dois objetivos. Evidentemente, tanto o escopo deste estudo quanto seu espaço permitem apenas uma breve apresentação dos mais comuns; versão estendida deste trabalho, a ser publicada como *Texto para Discussão*, os discutirá com mais detalhes.

Uma das maneiras mais intuitivas de fazê-lo é lançar mão dos chamados “experimentos quasi-naturais”: algo que torne a atribuição da intervenção (a definição se uma unidade será tratada ou controle) exógena, seja por um acontecimento histórico, por uma regra arbitrária de participação, ou mesmo por erros na implementação de alguma política.

Outra técnica bastante popular em pesquisa aplicada é o método de “diferenças-em-diferenças”. Como o nome sugere, essa técnica combina as duas comparações “ingênuas” discutidas anteriormente para identificar o efeito causal de uma intervenção. A hipótese central é que, na ausência da intervenção, as variáveis de interesse de tratados e controles teriam a mesma trajetória temporal – ou seja, o efeito das variáveis omitidas seria o mesmo nas duas unidades. Caso isso seja verdade, mesmo que o controle não seja um bom contrafactual para o tratado, a trajetória do controle será um bom contrafactual para a trajetória do tratado.

Um conjunto de métodos se baseia na hipótese de que a atribuição da intervenção depende apenas de características observáveis das unidades – os chamados métodos de pareamento.

Há ainda um conjunto de métodos que se baseiam na existência de particularidades no mecanismo de atribuição que introduza alguma exogeneidade na seleção, tais como o método de regressão com descontinuidade, que examina os indivíduos em torno de uma descontinuidade arbitrária na regra de elegibilidade de alguma política, que se supõe que sejam similares em diversas dimensões relevantes.

5 LIMITAÇÕES E CONSIDERAÇÕES FINAIS

Este artigo teve por propósito discutir a ideia de inferência causal, ressaltando a importância do conceito de contrafactual nessa tarefa, e chamando atenção para a inadequação de comparações intuitivas do tipo antes *versus* depois e afetados *versus* não afetados no exame de questões de natureza causal.

Embora meritória, a preocupação com a qualidade da inferência causal na avaliação de políticas públicas tem suas limitações, a começar pelo fato de que pode reduzir o escopo das perguntas de pesquisa que podem ser respondidas. Além do (óbvio) fato já apontado de que nem toda intervenção pode ou deve ser aleatorizada, esse problema pode surgir mesmo no caso de técnicas para dados observacionais – podendo levar a uma situação em que a busca por uma identificação adequada pode desincentivar a análise de problemas que, embora fundamentais, não seja possível obter evidência causal satisfatória, como apontado por Ruhm (2019), por exemplo.

Paralelamente a esse problema – técnica ditando a direção da pesquisa – pode-se vislumbrar outro, potencialmente mais arriscado, de as necessidades da avaliação de políticas ditarem seu desenho e implementação, ou o acesso a elas.

Além disso, há que se ter em mente que um insumo fundamental para uma boa inferência causal é a disponibilidade de dados adequados, e a mensuração adequada de fenômenos sociais muitas vezes pode ser uma tarefa ainda mais desafiadora. A utilização complementar de métodos qualitativos pode ser uma importante ferramenta nesse sentido.

Mesmo que todos esses desafios sejam circunvençionados, resta a questão da validade externa: como um *policymaker* pode ter segurança que conclusões alcançadas em outros contextos podem informar sua decisão? A esse respeito, Williams (2020) propõe um método arcabouço inovador para tratar a adaptação de políticas em contextos diferentes.

Em que pesem essas limitações, é essencial que formuladores, implementadores e analistas de políticas públicas sejam dotados do ferramental necessário para pensar de maneira adequada sobre questões de natureza causal – a começar pela mais fundamental de todas: “essa política funciona?”

REFERÊNCIAS

- HOLLAND, P. W. Statistics and causal inference. **Journal of the American Statistical Association**, v. 81, n. 396, p. 945-960, 1986. Disponível em: <https://doi.org/10.1080/01621459.1986.10478354>.
- IMBENS, G. W. Potential outcome and directed acyclic graph approaches to causality: relevance for empirical practice in economics. **Journal of Economic Literature**, v. 58, n. 4, p. 1129-1179, 2020. Disponível em: <https://doi.org/10.1257/jel.20191597>.

KING, G.; KEOHANE, R. O.; VERBA, S. **Designing social inquiry**: scientific inference in qualitative research. Princeton: Princeton University Press, 1994.

KOOPMANS, T. C. Measurement without theory. **The Review of Economics and Statistics**, v. 29, n. 3, p. 161, 1947. Disponível em: <https://doi.org/10.2307/1928627>.

MIZE, T. D.; MANAGO, B. The past, present, and future of experimental methods in the social sciences. **Social Science Research**, v. 108, p. 102799, 2022. Disponível em: <https://doi.org/10.1016/j.ssresearch.2022.102799>.

NEYMAN, J. On the application of probability theory to agricultural experiments. Tradução de Dorota M. Dabrowska e Terence P. Speed. **Statistical Science**, v. 5, n. 4, p. 465-472, 1990 [1923].

PEARL, J.; MACKENZIE, D. **The book of why**: the new science of cause and effect. 1st ed. [s.l.]: Basic Books, 2018.

RUBIN, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. **Journal of Educational Psychology**, v. 66, n. 5, p. 688-701, 1974. Disponível em: <https://doi.org/10.1037/h0037350>.

RUHM, C. J. Shackling the identification police? **Southern Economic Journal**, v. 85, n. 4, p. 1016-1026, 2019. Disponível em: <https://doi.org/10.1002/soej.12333>.

WILLIAMS, M. J. External validity and policy adaptation: from impact evaluation to policy design. **The World Bank Research Observer**, v. 35, n. 2, p. 158-191, 2020. Disponível em: <https://doi.org/10.1093/wbro/lky010>.