

Evidência: o que é, para que serve e como ir da que temos para a que queremos?

CÁTEDRA

Instituto Ayrton Senna
Núcleo Ciência para Educação

Ricardo Paes de Barros

Daniel D. Santos

Diana Coutinho

Camila M. M. Soares



Insper

Ciência e pesquisa certamente têm múltiplas finalidades e podem ser justificadas pela simples necessidade humana de compreender o mundo a sua volta. Mas, para que esta seja realmente importante, inquestionavelmente a ciência e a pesquisa precisam servir para mudar a vida das pessoas para melhor. Como afirma Howard White, CEO da Campbell Collaboration, em seu discurso de abertura da What Works Global Summit em setembro de 2016, “What is research good for? Absolutely nothing! Unless we can get it into policy/practice”.

Como iremos também argumentar mais adiante e é de conhecimento geral, do ponto de vista prático, enquanto os cientistas visam gerar evidência e os formuladores de políticas públicas visam conhecer essa evidência e tomar decisões dando-lhe a devida consideração. Assim, fica claro que a evidência é o elo central da conexão entre ciência e política pública e, dessa forma, é o que permite à ciência e à pesquisa alcançarem sua devida importância social. Mas o que é evidência?

O QUE É EVIDÊNCIA?

Segundo o Dicionário Houaiss evidência é a “constatação de uma verdade que não suscita qualquer dúvida, pela clareza e distinção com que se apresenta”, a “qualidade ou caráter do que é evidente, do que não dá margem à dúvida”. Já segundo o Dicionário Aurélio, evidência é o “caráter de objeto de conhecimento que não comporta nenhuma dúvida quanto à sua verdade ou falsidade”.

A característica comum a todas essas formas de descrever evidência é a sua relação com “certeza” e “ausência de dúvida”. Assim, evidência, quando relacionada à certeza, nunca pode ser “relativa”, deve sempre ser “absoluta”. Dessa forma, algo que dá margem a dúvida nunca poderá ser classificado como evidência. A relação da evidência com a certeza é binária e portanto, de nenhuma forma, contínua. Como veremos abaixo, o que pode ser tratado num contínuo não é a evidência per se, mas em que medida a interpretação dada à evidência encontrada – um fato, resultado ou ocorrência (uma correlação, por exemplo) – constitui-se em uma explicação plausível de algo de particular interesse (uma relação causal, por exemplo).

EXISTE ALGUMA DIFERENÇA ENTRE EVIDÊNCIA E EVIDÊNCIA CIENTÍFICA?

No contexto acadêmico, há a definição de evidência científica, compreendida como “os resultados de pesquisas objetivas e científicas, obtidas por meio de procedimentos que incorporaram critérios de validade, minimizando-se o grau de viés” (DE-LA-TORRE-UGARTE-GUANILO, M.; TAKAHASHI, R.; BERTOLOZZI, M. 2011). Além disso, para que uma evidência seja científica, é preciso que a pesquisa que deu origem a esta evidência obedeça a quatro critérios, de viabilidade, adequação, significância (pertinência) e eficácia, conhecidos como FAME (Feasibility, Appropriateness, Meaningfulness, Effectiveness) (DE-LA-TORRE-UGARTE-GUANILO, M.; TAKAHASHI, R.; BERTOLOZZI, M. 2011).

Pode-se, assim, dizer que evidência científica é algo mais específico que evidência e possui uma definição mais restrita. Nesse texto, não adotamos a definição de evidência científica apresentada acima, optamos por entender evidência como no cartesianismo: “constatação de uma verdade que não suscita qualquer dúvida” (Dicionário Housaiss).

Importante destacar que não há nenhuma restrição ao formato da evidência. Ou seja, a evidência pode ter origem em análises quantitativas ou qualitativas, podendo assumir a forma de um texto, um número, uma foto ou um depoimento; nada disso define ou impede que algo traga certeza e, portanto, todas essas podem ser formas válidas de evidência.

Ainda, quanto à definição de evidência, é importante ressaltar sua relação com a lógica, que será central para toda a nossa argumentação neste documento. Como a certeza é preservada pela lógica (própria definição de lógica), tudo que decorre – de forma lógica – de uma evidência, é necessariamente uma nova evidência.

O QUE SIGNIFICA TOMAR DECISÕES BASEADAS EM EVIDÊNCIAS? E POR QUE ADVOGAMOS A FAVOR DISSO?

Diante da necessidade de desenhar, aperfeiçoar, ou simplesmente implementar uma ação, programa ou política, um gestor público baseia suas decisões necessariamente em quatro elementos:

1. O grupo de interesse que pode estar representando
2. Seus valores ou os daqueles que representa
3. A evidência disponível
4. Sua interpretação subjetiva da evidência disponível

É certamente indesejável que dois gestores, que representem os mesmos interesses, compartilhem dos mesmos valores, e possuam interpretações similares da evidência disponível, tomem decisões diferentes apenas por terem tido acesso diferenciado à evidência disponível. Nesse caso, as decisões irão diferir apenas porque um grupo teve acesso a determinadas evidências e, com isso, sabe com certeza algumas coisas, ao passo que o outro grupo inferiu (interpretou) incorretamente a partir da limitada evidência a que teve acesso. Assim, é fundamental garantir acesso pleno à toda evidência existente a todos os gestores públicos.

É evidente que são poucas as decisões de política pública tomadas exclusivamente com base em evidência. Afinal, via de regra, não há evidências suficientes para solucionar um problema ou embasar completamente uma decisão. Invariavelmente, o gestor precisa combinar a evidência disponível com interpretações da evidência para tomar suas decisões. Assim, o avanço da ciência é importante para aumentar a quantidade de evidências disponíveis e reduzir a necessidade do gestor público tomar decisões com base em interpretações.

Indiretamente, a ampliação do leque de evidências disponíveis restringe a importância dos interesses e dos valores na tomada de decisões. Isso ocorre pois interesses e valores podem afetar a interpretação da evidência, mas nunca afetam a evidência em si.

Importante também ressaltar que a interpretação do cientista não é, per se, superior à do gestor público. O fato de a interpretação do gestor ser, via de regra, baseada em sua experiência e intuição, enquanto a dos cientistas é baseada em teorias mais ou menos comprovadas, não torna a interpretação do cientista em nenhum grau superior à interpretação do gestor.

TENTATIVAS DE TRANSFORMAR OBSERVAÇÕES EM EVIDÊNCIA

É incontestável que observações e mensurações são evidência. A questão é elas são evidência do que. Tipicamente, temos evidência de uma coisa (que duas variáveis estão correlacionadas, por exemplo) quando o que realmente queremos é evidência de outra coisa (que um fator causa outro). Muito raramente é possível “constatar uma verdade desejada de uma forma que não suscite qualquer dúvida”, porque, simplesmente e infelizmente, não existe qualquer tendência de que a evidência disponível case bem com as verdades que desejamos estabelecer. Assim, por nossa definição de evidência baseada no Dicionário Housais, a evidência é um objeto raro. Em nosso mundo ordinário, contamos com um conjunto de observações e mensurações (evidência) que são compatíveis com a verdade que desejamos estabelecer, mas que não levam, via uma série de passos lógicos, à demonstração dessa verdade. Nesse sentido, temos, em geral, muita evidência, mas muito pouca evidência relativa ao que efetivamente queremos. Por esse motivo, tentamos sempre transformar a evidência que temos na que queremos. Uma forma de fazer essa ponte é pelo uso de hipóteses. Em muitos casos, a evidência que temos, somada a um conjunto de hipóteses, asseguram com certeza a verdade que se deseja estabelecer (evidência desejada).

Em outras palavras, na prática, para se obter a evidência desejada, combinamos a evidência existente com um conjunto conveniente de hipóteses. Assim, embora não se possa relativizar o que vem a ser a “evidência desejada”, o quanto a “evidência existente” pode ser considerada equivalente à “evidência desejada” depende integralmente do quanto as hipóteses que fazem essa ponte sejam consideradas válidas. É evidente que uma hipótese ou é verdadeira ou é falsa e, portanto, ou as hipóteses consideradas levam a “evidência existente” à “evidência desejada” ou simplesmente não levam, porque ao menos uma delas é falsa. Mas, como a veracidade dessas hipóteses não é conhecida, estabelecemos um julgamento subjetivo sobre sua validade e daí decorre a “relatividade” da equivalência entre a “evidência existente” e a “evidência desejada”.

Essa relatividade das hipóteses que fazem a ponte entre a “evidência existente” e a “evidência desejada” pode certamente ser hierarquizada de variadas formas. Apresentamos a seguir uma hierarquização baseada no grau em que essas hipóteses são explicitadas, no grau em que se baseiam em teorias com maior ou menor comprovação e no grau em que são percebidas como plausíveis.

Não é difícil de perceber (embora possa ser difícil de demonstrar) que, caso uma “evidência existente” seja compatível com uma “evidência desejada”, então existe um conjunto de

hipóteses que, se válidas, permitem passar via passos lógicos da “evidência existente” à “evidência desejada”. Assim, numa ladeira hierárquica, podemos dizer que a etapa mais básica é estabelecer compatibilidade entre a “evidência existente” e a “evidência desejada”.

Desse nível básico, seguem outros cinco, formando a seguinte hierarquia:

NÍVEL 1: A “evidência existente” é compatível com uma “evidência desejada” e, portanto, existe um conjunto de hipóteses que, se válidas, permitem passar via passos lógicos da “evidência existente” à “evidência desejada”. Não se conhece, neste nível, quais seriam essas hipóteses.

NÍVEL 2: A “evidência existente” é compatível com uma “evidência desejada” e o conjunto de hipóteses que, se válidas, permitem passar via passos lógicos da “evidência existente” à “evidência desejada” é conhecido.

NÍVEL 3: A “evidência existente” é compatível com uma “evidência desejada”, o conjunto de hipóteses que, se válidas, permitem passar via passos lógicos da “evidência existente” à “evidência desejada” é conhecido e baseado parcialmente em alguma teoria ou em combinações coerentes de várias.

NÍVEL 4: A “evidência existente” é compatível com uma “evidência desejada”, o conjunto de hipóteses que, se válidas, permitem passar via passos lógicos da “evidência existente” à “evidência desejada” é conhecido e baseado parcialmente em alguma teoria ou em combinações coerentes de várias e parcialmente em hipótese adicionais necessárias (que denominamos de hipóteses operacionais) que são consideradas plausíveis.

NÍVEL 5: A “evidência existente” é compatível com uma “evidência desejada”, o conjunto de hipóteses que, se válidas, permitem passar via passos lógicos da “evidência existente” à “evidência desejada” é conhecido e baseado integralmente em alguma teoria ou em combinações coerentes de várias.

NÍVEL 6: A “evidência existente” é compatível com uma “evidência desejada”, o conjunto de hipóteses que, se válidas, permitem passar via passos lógicos da “evidência existente” à “evidência desejada” é conhecido e baseado integralmente em hipóteses consideradas plausíveis.

QUAIS TIPOS DE EVIDÊNCIA O GESTOR USA?

Sete tipos de evidências são usadas pelo gestor público na tomada de decisões:

- 1- Evidências que permitam identificar progressos e avanços, fracassos e retrocessos ou a detecção de novas tendências.
- 2- Evidências que possibilitem identificar os determinantes tanto dos progressos, quanto dos fracassos e das novas tendências.
- 3- Evidências que possibilitem identificar as consequências dos progressos, fracassos e novas tendências.
- 4- Evidências da capacidade de mobilização de recursos da sociedade para determinadas causas.
- 5- Evidências da eficiência com que recursos mobilizados são utilizados.
- 6- Evidências sobre adequação do processo de seleção dos beneficiários, ou seja, se estão sendo atendidos aqueles que deveriam ser atendidos.

7- Evidência de impacto baseada em avaliações ex-ante ou ex-post da ação, programa ou política em foco ou evidência de impacto derivada de avaliações de programas similares (meta-análise).

O primeiro tipo de evidência é parte de qualquer bom sistema de monitoramento e não envolve relações causais. Esse tipo de evidência indica para o gestor público qual é a situação de seu município, estado ou país.

Já o segundo, o terceiro e o sétimo tipos de evidência são utilizados para que o gestor possa saber como (2 e 7) e por que (3) modificar a situação em que se encontra – identificada com o primeiros tipos de evidência. Para tal, o gestor precisa, em geral, entender a causa de uma situação e, portanto, necessita de evidências causais.

Por que o gestor público precisa de evidência causal?

Uma vez identificado um resultado que precisa ser modificado, o gestor público necessita de evidências sobre como realizar essa modificação. Dessa forma, precisa identificar e avaliar a relativa importância dos determinantes (os fatores que causam) do resultado que se deseja modificar. Quando o determinante é uma ação, denominamos de magnitude do impacto o seu efeito sobre o resultado de interesse. Assim, a magnitude do impacto de uma ação aponta para quão efetivamente essa ação leva (“causa”) a melhorias no resultado de interesse. Evidências causais não se referem, portanto, a evidências de associações, mas, sim, ao isolamento do impacto individual de diferentes ações ou determinantes.

COMO USUALMENTE PARTIMOS DA EVIDÊNCIA DE CORRELAÇÃO E JULGAMOS A FAVOR DE CAUSALIDADE?

O que é correlação ou poder preditivo?

É em essência um exercício empírico de adivinhação. Imagine que preciso estabelecer minha melhor aposta no valor de uma variável Y, por exemplo a altura de uma pessoa. Se ninguém me disser mais nada sobre essa pessoa, talvez meu melhor palpite seja a altura média (ou mediana) da população. Mas se eu souber que a pessoa tem 10 anos de idade, eu certamente mudaria meu palpite (talvez para a média daqueles que têm 10 anos de idade).

No caso da idade, talvez fosse pensar que a razão para eu mudar meu palpite seja que mudanças de idade de fato causam (ao menos até certo patamar) a mudança de altura. Eventualmente a informação adicional poderia ser também a de que o nome da pessoa é Ariosvaldo, e apesar de o nome da pessoa não ser causador da altura, sabemos nesse caso que uma terceira variável, sexo, causa ambos o nome ser Ariosvaldo e minha adivinhação da altura ser diferente da altura média da população. Pelo fato de o nome ser este, sei que o indivíduo é homem e que os homens são em média mais altos do que as mulheres, induzindo a subir meu palpite em relação à situação em que não sabíamos nada a respeito da pessoa. Poderíamos ir adiante e imaginar que este nome é incomum nas coortes mais novas, o que

talvez nos fizesse também ajustar nosso palpite para baixo, pois sabemos que gerações mais antigas são em média mais baixas do que as mais novas.

Por outro lado, algumas informações podem não ajudar em nada meu exercício de adivinhação. Se nos fosse informado de que a pessoa é auxiliar de escritório, não creio que isto em si mudasse nosso palpite sobre sua altura. Para que a informação adicional (chamemos de X) nos faça melhorar nossa opinião sobre Y, é necessário que X e Y estejam estatisticamente relacionadas (ou que acreditemos que o estão). A forma mais comum de se estabelecer um co-movimento entre duas variáveis é a correlação, que captura uma relação linear sistemática entre duas variáveis. Por isso o termo “poder preditivo” aparece frequentemente associado a esta estatística. Em políticas públicas, especialmente em monitoramento, com frequência precisamos adivinhar quem são os indivíduos que mais precisam de um serviço.

Qual a relação entre correlação e causalidade?

É comum ouvirmos acadêmicos dizerem que correlação não é causalidade, colocando estas duas ideias em campos opostos. Na verdade, a correlação é em geral um indício de um fenômeno causal, ainda que seja difícil dizer de antemão qual. Duas variáveis (X e Y) não começam a andar juntas espontaneamente na natureza. Se constatamos que a correlação entre ambas é não nula, uma potencial explicação é a de que X causa Y. Alternativamente, se for Y a causar X também perceberíamos uma correlação não nula, e se uma terceira variável, Z, causar ambas X e Y; ao fim do dia possivelmente teríamos correlação não nula entre X e Y, ainda que o fenômeno causal por trás dessa relação não seja direto. Apenas como comentário adicional, essa terceira variável Z pode ser tanto algo que concretamente pertence (ou deveria pertencer) a uma teoria de mudança de Y e X, quanto um artefato dos dados. Use como exemplo as medidas psicológicas ipsativas. Se uma característica dos meus dados é que apenas observo X e Y entre indivíduos para os quais a soma de X e Y é 100, sei que um indivíduo com X elevado precisa por construção ter Y baixo e vice-versa, de modo que nos dados apareceria uma falsa correlação negativa entre estas variáveis, que tem como fenômeno causal subjacente a forma como os dados foram construídos (problema semelhante é de seletividade, descrito por James Heckman).

Em termos de evidência científica, quase sempre se acredita que quando há correlação (e não se sabe a priori qual o fenômeno causal que está por trás), a explicação causal “X causa Y” é ao menos um dos candidatos a ser a resposta. Aceitar esta opção é semelhante a tomar uma decisão de investimento, no sentido de que a solução causal tem chances de ser verdadeira, e nesse caso pode levar a importantes indicações decisórias para stakeholders, mas sem negligenciar que com alguma probabilidade pode não ser a resposta desejada. “Investir” nessa explicação depende fundamentalmente de como estimamos intimamente a chance de que seja a solução correta, bem como das consequências em torno de recomendar esta solução a outrem sem que seja correta. Note que, ainda que teoricamente seja possível que haja efeito causal de X para Y sem que se note concomitantemente uma correlação não-nula, a maioria das pessoas espera que se X de fato causa Y então alguma associação

estatística deveria ser notada entre estas variáveis, de modo que, se por um lado correlação não é necessariamente atribuível à existência de um efeito causal subjacente, por outro lado a ausência de correlação quase sempre é interpretada como evidência de ausência de efeito causal (e desencorajadora de esforços mais intensos de investigação para determinar se há ou não causalidade).

Quando e como alguém se convence de que a correlação (ou qualquer outro parâmetro estatístico que capture associação sistemática entre X e Y) estimada pode ser atribuída a um efeito causal?

As estratégias de pesquisa quase sempre tentam nos convencer de que a associação se deve a um efeito causal pela eliminação das explicações alternativas. O caminho mais comum (não necessariamente mais rigoroso) é a tentativa de controlar para fatores confundidores. Uma explicação alternativa para a associação entre X e Y costuma ser a de que terceiras variáveis, inicialmente omitidas da análise, ao mesmo tempo influenciam X e Y , e, portanto, acabam por fazer que estas duas variáveis caminhem juntas. Uma análise inicial, por exemplo, poderia sugerir que o fato de que brancos ganham mais do que negros no mercado de trabalho é evidência de que haja discriminação (em economia definida como remuneração diferencial de indivíduos associada a características não-produtivas dos mesmos). Uma explicação alternativa seria a de que brancos em média estudam mais do que negros, e é esta variável omitida que efetivamente induz a uma relação sistemática entre etnia e salários. Um esforço de pesquisa pode então comparar brancos e negros com mesmo nível educacional, e verificar se agora, no interior de grupos educacionalmente homogêneos, ainda restam diferenças salariais. Se restarem, descartamos a explicação alternativa (educação omitida) e mantemos a discriminação no páreo. Caso no interior de grupos educacionalmente homogêneos já não haja diferenças salariais, a maioria dos pesquisadores aceitará a explicação de que era a educação que conduzia as diferenças salariais originalmente verificadas, e não a cor da pele. Bases de dados mais ricas permitem manter fixas mais características, e avaliar o efeito líquido de X sobre Y .

O problema com a estratégia acima é o de que, por mais rica que seja nossa base de dados (e nem sempre o é), apenas conseguimos fazer comparações entre grupos homogêneos em características mensuráveis. Seguindo o exemplo acima, e imaginando que mesmo em grupos educacionalmente homogêneos restassem diferenças de salários por etnia, ainda assim um observador mais crítico poderia dizer que as diferenças seriam alternativamente atribuíveis a, digamos, motivação intrínseca distinta entre estes grupos (talvez escorado em uma teoria social que dissesse que o acúmulo de frustrações anteriores à entrada no mercado de trabalho produzisse sequelas na crença individual de que se pode vencer na vida). Como esta motivação é difícil de medir, a dúvida permaneceria, e “investidores” continuariam mantendo a explicação causal como incerta (ou “arriscada”).

Para lidar com esta dificuldade, uma estratégia possível é explorar a variação longitudinal de indicadores. Vamos mudar nosso exemplo para entender como funciona. Suponha que se

queira verificar se uma fórmula, “vitamina X” afeta a temperatura corporal, Y, dos indivíduos. A comparação direta de temperatura entre indivíduos com elevada concentração natural de vitamina X com os que possuem naturalmente baixos níveis dessa substância poderia não ser causal se, por exemplo, as mesmas pessoas que possuem mais vitamina X também possuem mais vitamina A, iodo, e outros micronutrientes. Sendo verdade, e assim como antes, as diferenças de temperatura entre os dois grupos talvez fosse resultado de diferenças nas concentrações naturais destes outros micronutrientes, e não da vitamina X. Mesmo após medir uma coleção destes outros micronutrientes, há suspeita de que substâncias não-mensuráveis ficaram de fora e seriam a verdadeira fonte de diferenças de temperatura entre indivíduos. Mas poderíamos ter agora dados longitudinais que nos permitissem relacionar variações dos níveis de vitamina X com variações dos níveis de temperatura (ao invés de relacionar níveis destas variáveis). Neste caso, não mais estaríamos comparando indivíduos distintos em um mesmo ponto do tempo, mas sim um mesmo indivíduo em pontos distintos do tempo, e verificando se em instantes em que há concentrações particularmente elevadas de vitamina X em seu corpo, também se observa redução da temperatura e vice-versa. Para que esta estratégia seja verdadeira, é preciso que variações na concentração de vitamina X não venham acompanhada de variações de outros micronutrientes (especialmente os difíceis de observar e conseqüentemente manter fixo). Note que apesar de, em termos de mecanismo, termos basicamente uma troca de níveis por variações na forma como procedemos com as estimativas, existem várias teorias (genéticas e ambientais) que levantam a suspeita de que as mesmas pessoas que têm grande concentração de uma substância no corpo também teriam grandes concentrações de outras substâncias, mas há bem menos teorias que indicariam razões para suspeitarmos que nos mesmos instantes em que a concentração de vitamina X aumenta também o aumentam as concentrações de outras substâncias, e por isso muitas vezes a comparação longitudinal torna-se mais convincente do que a comparação “natural”.

A terceira grande estratégia para determinar causalidade (ou eliminar explicações alternativas) é a separação aleatória entre indivíduos que receberão ou não uma injeção de vitamina X. Nesse caso, o interesse é menor em estimar, para cada grupo homogêneo em determinado grupo de características, o efeito de X sobre Y, mas em compensação estimar de modo altamente convincente o efeito de um acréscimo de X sobre a média de Y. A ideia agora é a de que, logo após separar aleatoriamente dois grupos de uma população original, temos duas subpopulações que herdaram em princípio mesmas características médias de todas as dimensões que poderiam afetar Y. Qualquer mudança em Y após uma alteração induzida em X em um dos grupos deveria, portanto, ser atribuível a este evento, uma vez que os dois grupos eram (estatisticamente) idênticos ex-ante. Explicações alternativas para esta relação estatística (aumento na média de Y concomitante ao aumento na média de X) são difíceis de imaginar. A possibilidade, por exemplo, de causalidade reversa (que fosse Y a causar X do que o reverso) fica excluída pela antecedência temporal da mudança em X neste caso. A possibilidade de que uma terceira variável Z fosse o verdadeiro causador de mudanças em X e Y simultaneamente tampouco é fácil de imaginar, uma vez que a “injeção de vitamina X” é mecanismo relativamente ao controle do implementador (opa! A possibilidade de efeito

placebo não está em princípio excluída e se enquadra nessa possibilidade! Cabe ao pesquisador tomar precauções quanto a isso).

Uma limitação forte da aleatorização, contudo, é a dificuldade de associar suas conclusões com uma teoria de mudança. Em primeiro lugar, o efeito medido é em princípio o efeito total de uma mudança em X sobre Y, em contraste com o efeito direto. Se por exemplo parte do efeito de X sobre Y ocorre porque X influencia uma segunda variável, W, que por sua vez afeta Y, não poderemos separar trivialmente este efeito do efeito direto “X causa Y”. Modelos de mediação e moderação que declaram conseguir separar mecanismos sofrem das mesmas potenciais críticas mencionadas anteriormente. Diga-se de passagem essa limitação anteriormente era resolvida de maneira potencialmente igualmente arbitrária pela forma como o pesquisador inclui fatores confundidores em sua estimação, mas estimações que seguem as estratégias anteriores e ainda assim permanecem convincentes costumam ter por trás teorias mais sólidas que justifiquem precisamente o modo pelo qual tais variáveis deveriam entrar no problema (o que significa que, dentre trabalhos acadêmicos bem publicados, é mais comum encontrar melhores ideias sobre como desvendar mecanismos nas estratégias de convencimento de causalidade inicialmente apresentadas do que dentre artigos que usam aleatorização). Em segundo lugar, estudos aleatorizados medem precisamente o efeito médio de tratamento, mas são pouco informativos sobre efeitos individuais. Quando se deseja extrapolar por exemplo as conclusões obtidas em determinado contexto para outras realidades, o natural seria primeiro verificar a diferença na composição de ambas as populações, e em seguida agregar efeitos individuais segundo a composição da população para a qual se deseja extrapolar as conclusões. Para dar tal passo com frequência se necessita recorrer a métodos análogos aos iniciais, perdendo as vantagens da manipulação controlada do exercício aleatorizado. Finalmente, e ao contrário dos exercícios de aleatorização geralmente feitos em outras áreas do conhecimento, a população original sobre a qual se aleatoriza costuma ser bastante heterogênea nas características envolvidas nos processos causais em problemas de ciências sociais aplicadas. Quando se testa um princípio ativo como potencial cura para uma doença, por exemplo, a população original compartilha sintomas semelhantes e a presença de eventual agente causador da doença no organismo, e são estas as principais características envolvidas no mecanismo causal de eliminação do agente pelo princípio ativo. Consequentemente, é mais fácil acreditar que conclusões obtidas em uma amostra podem ser extrapoladas, e a relativa facilidade de replicar o exercício em outros contextos pode inclusive demonstrar (como em geral demonstra) que efeitos bastante semelhantes são esperados se o princípio ativo for ingerido qualquer que seja a cultura ou ambiente em que isso seja feito. Agora avalie o caso de uma intervenção educacional, digamos, acesso a creches de determinada qualidade (imagine que a intervenção seja idêntica em diferentes contextos, o que dificilmente é o caso). Um exercício aleatorizado rigoroso feito, por exemplo, na Suécia, pode encontrar um efeito nulo ao passo que sua replicação no Iraque pode encontrar um efeito fortemente positivo. A explicação é que o grupo de controle, ao contrário do descrito no caso do princípio ativo, é muito diferente nestes dois casos. No caso sueco, apenas se detectaria um efeito positivo se a creche acrescentasse algo no topo do estágio de desenvolvimento que uma criança sueca naturalmente teria com as oportunidades

que possui, ao passo que no Iraque o grupo de controle possivelmente tem menos oportunidades e torna-se menos difícil obter impactos positivos. Com isso, percebe-se que a estratégia de aleatorização permite concluir, para o contexto em que foi feita, se houve progresso no topo do que teria ocorrido na ausência da aleatorização, mas não permite em princípio concluir se creches fazem bem para as crianças.

O papel da teoria

Como foi dito, a ciência não consegue confirmar uma explicação sobre o mundo, mas sim eliminar explicações alternativas que não se conformam aos dados e outros elementos da realidade observáveis. No processo de eliminação de explicações alternativas, a comunidade acadêmica não conta apenas com refinamentos de exercícios empíricos, seja através de obtenção de dados melhores, seja através de sofisticação da técnica. Os avanços teóricos, através da eliminação lógica de possibilidades implausíveis, desempenha papel central neste processo, e uma mesma correlação pode ter diferentes graus de convencimento sobre terceiros se conectada a uma teoria sólida e que, em outras partes, esteja empiricamente respaldada.

Mais ainda do que no “simples” desafio de ter uma estimativa causal, a teoria é indispensável se a intenção é dar um passo adiante e desvendar mecanismos, uma vez que raramente é prático ou financiável a estratégia de sucessivas randomizações para determinar dosagens, complementaridades entre causas, etc.

Finalmente, a teoria é também fundamental para convencer as pessoas a respeito de possibilidades de extrapolação de conclusões obtidas em um contexto para outros contextos, pois é preciso para isso ou (i) acreditar que a forma como X se relaciona com Y não depende em nada do contexto, ou (ii) descrever exatamente como depende do contexto para que a partir de um ponto inicial se saiba como extrapolar. A amostra original em que foi feito o exercício não permite obter tais conclusões apenas com uso dos próprios dados.

Conclusões

O uso de evidência para fazer recomendações de política é análogo ao processo de desenvolvimento da ciência. Cada fato empírico obtido a partir dos dados (ou outros elementos da realidade) possui em princípio múltiplas interpretações, e o progresso é feito através da eliminação de explicações implausíveis até que as pessoas se convençam de que a explicação causal é a mais provável e passem a apostar nela. A cada passo, a incerteza em torno da interpretação causal como verdadeira explicação se reduz, e menor é o risco de errar ao se recomendar uma intervenção que altere o nível de X como forma eficaz de afetar Y. O “investimento” nesta orientação torna-se, portanto, mais rentável e menos arriscado conforme a certeza em torno da causalidade aumenta. A certeza absoluta em torno da causalidade é sonho antigo, mas inalcançável em ciências sociais (ou alcançável em situações tão específicas que quase tornam-se irrelevantes). O montante de energia, recursos e tempo investidos deve depender, caso a caso, do grau de certeza que o “investidor” forma em torno do problema. O principal papel da evidência na tomada de decisão ainda é o de permitir que

eventualmente se mude de opinião e pare de investir, caso nova informação chegue que altere a expectativa de retorno (muito mais do que o de produzir um nível absoluto de certeza que garanta a rentabilidade). Um complicador deste processo são situações em que, pela natureza do processo decisório, sabe-se que o tomador de decisão precisará fazer algo com ou sem a orientação da ciência. Nestes casos, um conselheiro que se preocupe com o bem-estar coletivo deveria incluir dentre os benefícios de seu conselho os potenciais erros que uma decisão completamente cega produziria, e possivelmente o grau de certeza mínimo necessário para valer a pena a recomendação diminui se comparado à situação em que o tomador de decisão pode esperar pelo convencimento dos cientistas sobre o tema.