

As implicações metodológicas da Revolução da Credibilidade e Inferência Bayesiana para a pesquisa qualitativa causal

Manoel Galino

2024-06-30

Abstract

Esta nota de pesquisa argumenta que os avanços recentes na metodologia de inferência causal — notadamente a Revolução da Credibilidade nos métodos quantitativos e o desenvolvimento da inferência Bayesiana — oferecem um novo enquadramento para avaliar o rigor e a validade da pesquisa qualitativa em ciência política. Em vez de assumir que métodos qualitativos são inerentemente limitados em sua capacidade de generalização, propomos uma distinção mais precisa entre identificação causal e inferência estatística. Com base em dois desenvolvimentos recentes — o process tracing Bayesiano e o modelo de queries causais — mostramos que estudos de pequeno-n podem produzir inferências causais robustas, desde que as suposições de identificação sejam plausíveis, transparentes e avaliáveis. O objetivo da nota é sistematizar, de forma acessível e didática, os desafios e potenciais desses métodos, contribuindo para sua aplicação crítica no ensino e na pesquisa em ciência política.

1 Introdução

A pesquisa qualitativa em ciência política voltada ao estudo da causalidade, quando contrastada com o “template” quantitativo, é frequentemente percebida como carente de rigor metodológico e de capacidade para isolar efeitos causais. Essa concepção encontra respaldo em trabalhos clássicos, como o de King, Keohane, and Verba (1994) (KKV), que propôs a unificação da lógica da pesquisa em ciências sociais sob um paradigma quantitativo. Métodos como a análise comparativa de casos (Rihoux and Ragin 2009; Ragin 2014) e process tracing (Collier 2011; Bennett and Checkel 2015) são frequentemente considerados menos robustos que abordagens quantitativas precisamente porque suas inferências seriam restritas a contextos específicos e limitados por amostras pequenas Seawright (2018).

Em um artigo em que aponta as falhas do método comparativo qualitativo, Seawright (2018) elenca as críticas principais: para Sekhon (2004), de um ponto de vista probabilístico, sempre haverá explicações alternativas que não poderão ser eliminadas com estudos qualitativos; George and Bennett (2005) argumentam que na prática estudos comparados falham em satisfazer os requerimentos causais necessários; no livro *Rethinking Social Inquiry*, Brady and Collier (2010), juntamente com Seawright, argumentam que estudos comparados são inferencialmente fracos e, portanto, “inferentially as a weaker and problem-laden equivalent of statistical analysis” (Seawright 2018, 8). E arremata: “a potential-outcomes formulation makes evident that qualitative comparisons are exceptionally weak tools for causal inference” (ibid. p.8).

Por outro lado, tem havido um renovado interesse em refinar e desenvolver essas metodologias qualitativas por meio da utilização dos paradigmas causais de resultados potenciais e Directed Acyclic Graphs (DAGs), do qual é testemunho as recentes edições da *Qualitative and Multi-Method Research*, revista da seção de Qualitative and Multi-Method Research da American Political Science Association. Outro desenvolvimento recente também impactante é a adoção de métodos Bayesianos na pesquisa qualitativa.

O argumento desta nota de pesquisa é que o template quantitativo ao qual boa parte dos qualitativistas estão respondendo foi, em grande parte, rejeitado pelo próprio desenvolvimento metodológico da pesquisa quantitativa, conhecido como revolução da credibilidade.

Assim, apresentamos aqui o estado da arte do template quantitativo causal e como pode nos ajudar a reinterpretar os pontos fortes e fracos dos métodos comparados e process tracing. Nesse sentido, ambos os métodos podem seguir a mesma lógica de pesquisa, como sugerem King, Keohane, and Verba (1994) e Brady and Collier (2010), mas sob nova roupagem.

O estudo contribui para a literatura metodológica de quatro maneiras principais. Primeiro, esclarece a distinção entre identificação causal e inferência estatística, e por conseguinte, mostra que a incerteza chave nos estudos qualitativos causais é a da variabilidade nos resultados potenciais; e demonstra que a validade interna não é exclusiva dos métodos quantitativos. Segundo, contribui para uma literatura recente (Fairfield and Charman 2022; Humphreys and Jacobs 2023) que tem explorado como a estatística Bayesiana pode ser integrada ao process tracing e outras abordagens qualitativas para fortalecer a inferência causal em estudos de *n* pequeno. Terceiro, propõe uma reformulação das críticas metodológicas dirigidas à pesquisa qualitativa, sugerindo que a questão central não é a generalização, mas sim a credibilidade das suposições que sustentam a identificação causal e a quantificação da incerteza. Por fim, contribui para auxiliar no ensino e pesquisa de métodos qualitativos, de forma que suas forças e limitações possam ser mais bem apreciadas e não sejam desvalorizadas na pesquisa e ensino de ciência política.

O restante desta nota de pesquisa está organizada como segue: na próxima seção, discutimos a recepção da literatura qualitativa à crítica feita pelo template quantitativo. Em seguida, apresentamos a revolução da credibilidade e então como fazer inferências em amostras pequenas. Na quarta seção, introduzimos duas abordagens recentes que ilustram abordagens causais que enfrentam os desafios quantitativos. Por fim, discutimos questões de generalização e apresentamos as conclusões finais.

1.1 A recepção qualitativa ao template quantitativo

A visão proposta por KKV, como sabemos, foi alvo de vários debates e críticas, o que resultou no livro já clássico *Rethinking Social Inquiry* (Brady and Collier 2010). As críticas de KKV aos problemas de viés de seleção na seleção de casos foram largamente superados (Seawright 2018). Entretanto, grande parte das respostas dentro da metodologia qualitativa manteve

a suposição de que de fato a pesquisa qualitativa é limitada em termos de generalização (Amorim and Rodriguez 2016; Blatter and Haverland 2012; Brady and Collier 2010; George and Bennett 2005; Sekhon 2004). O foco dos contra argumentos passou por criticar: i) o custo de aumentar o tamanho da amostra e a consequente perda de contexto, ii) em defender a necessidade de diferenciar conceitualmente “data-observations” e observações de processos causais em “within case analysis” e iii) em defender que os objetivos são diferentes na pesquisa qualitativa (Brady, Collier, and Seawright 2010). Em última instância, no entanto, concedem capacidade de generalização para estudos quantitativos e a suposta necessidade de aumentar o tamanho da amostra para inferência.

Mahoney e Goertz (Mahoney and Goertz 2006), em um artigo clássico sobre as duas culturas, argumenta que boa parte da divisão entre quantitativistas e qualitativistas quando se trata de causalidade se deve a uma concepção distinta sobre o objetivo da explicação causal. Para os primeiros, estudos causais de casos particulares não fariam sentido. Contudo, o desenvolvimento do método de controle sintético é um contra-exemplo claro (Abadie, Diamond, and Hainmueller 2015), pois se trata justamente de estudar o efeito causal de um caso particular.

Outra consequência das crítica de KKV são a aceitação, por parte de pesquisadores qualitativos interessados em argumentos causais, é a ideia de que a comparação controlada não é capaz, em geral, de evitar o problema do viés de variável omitida Slater and Ziblatt (2013). Mais recentemente pesquisadores têm advogado a solução defendida por Slater and Ziblatt (2013): o método comparativo controlado permite separar entre hipóteses rivais principais, em vez de estimar o efeito causal de uma explicação específica. No entanto, não é claro em que isso permite superar o problema de viés de variável omitida. No presente trabalho, discutimos essa questão à luz de desenvolvimentos recentes nas metodologias quantitativas, em particular a ideia de esparsividade induzida por prioris. Antes disso, expliquemos em mais detalhes a revolução da credibilidade.

2 Revolução da Credibilidade

A revolução da credibilidade na economia (Forozish 2024; Goldsmith-Pinkham 2024; Angrist and Pischke 2010) impactou várias disciplinas que lidam com dados observacionais (Lundberg,

Johnson, and Stewart 2021; Libman 2023; Glied 2021), incluindo a ciência política (Samii 2016; Keele 2015b, 2015a; Grimmer 2015). Um dos trabalhos pioneiros que estabeleceram a agenda para essa revolução é o artigo de Leamer (1983), “Let’s take the con out of econometrics”, onde o autor critica a prática padrão de especificações econométricas múltiplas para controlar vieses de variável omitida, argumentando que tais práticas não produzem resultados confiáveis.

Respondendo ao desafio de Leamer, vários autores desenvolveram desenhos de pesquisa com suposições mais críveis e transparentes (Card 2022; Angrist and Pischke 2010, 2009; Angrist and Krueger 1991; Card and Krueger 1994). Esses estudos buscaram assegurar que os efeitos causais fossem reais e se aproximassem dos resultados de experimentos aleatórios controlados, elevando o padrão de inferência causal (Rubin 1974).

2.1 Identificação Causal

Identificação causal refere-se à capacidade de um desenho de pesquisa estimar um estimando específico, como o efeito causal médio (ATE - Average Treatment Effect) ou o efeito do tratamento sobre os tratados (ATT - Average Treatment on the Treated). A identificação não depende do tamanho da amostra, mas sim do desenho da pesquisa e de suposições críveis. Tipicamente, definimos causalidade para a unidade ou indivíduo i como sendo a diferença entre dois resultados potenciais (no caso de um tratamento binário): $Y_i(1) - Y_i(0)$, em que $Y_i(1)$ indica o resultado potencial para o indivíduo se recebeu o tratamento, $Y_i(0)$ se não recebeu (grupo controle).

É importante destacar que essa noção de causalidade **não** requer que a noção de causalidade seja estocástica. Há pelo menos três visões ou interpretações sobre de onde vem a aleatoriedade, quais sejam: i) aleatoriedade da amostra em relação à população; ii) aleatoriedade do tratamento ou controle; e iii) aleatoriedades dos resultados potenciais. Nas perspectivas i e ii, os resultados potenciais são fixos. Além disso, na perspectiva 1, o tratamento também pode ser fixo e a relação causal, portanto, determinística. O que gera a aleatoriedade é o processo amostral. Na segunda ontologia, resultados potenciais também são fixos e a aleatoriedade decorre da atribuição aleatória de tratamento e controle. Isso significa que um estudo com

dados populacionais ainda estimará um efeito causal aleatório (pois a atribuição de tratamento e controle mudaria de estudo para estudo). Apenas na terceira perspectiva a relação causal é inerentemente aleatória, pois os próprios resultados potenciais são estocásticos.

O ATE é simplesmente a média, na população, dos efeitos causais individuais. Para que estimandos como o ATE sejam identificados, isto é, estimáveis, mesmo com quantidade infinita de dados, é necessário algumas suposições. A mais típica é conhecida como “ignorability” forte, cuja suposição central é que a atribuição do tratamento é independente dos resultados potenciais, ou seja, as unidades não selecionam o tratamento com base nos seus potenciais benefícios (ou malefícios).

2.1.1 INUS e SUIN

Na metodologia qualitativa, é comum utilizar noções de causalidade que envolvam a noção de condição necessária e suficiente (Mahoney 2008). Para nós, o importante é destacar que podemos perfeitamente reescrever causalidade na notação de resultados potenciais como causas necessárias ou suficientes (ou os mais complexos SUIN e INUS) (A. Jacobs 2022). Digamos que temos três variáveis binárias potencialmente causais X, W, Z . Para representar a ideia de que $Z = 1$ é uma condição suficiente, mas não necessária para Y , basta escrever os resultados potenciais como: $Y(0, 0, 1) = 1, Y(0, 1, 1) = 1, Y(1, 1, 1) = 1, Y(1, 0, 1) = 1$ e (digamos) $Y(0, 0, 0) = 0$. Embora um pouco mais complicado, podemos também descrever condição INUS com resultados potenciais. Consideremos a causa potencial adicional K . Se $Y = 1$ pode ocorrer por combinação de X e W ou Z e K , então a condição é INUS. Ou seja, $Y(X, W, 1, 1) = 1$ para quais valores de X e W , $Y(1, 1, Z, K) = 1$ para quais valores de Z e K , e (digamos) $Y(1, 0, 1, 0) = 0$. Nesse exemplo, não há causa necessária para $Y = 1$, apenas potenciais causas suficientes em combinação. De maneira similar podemos estender o modelo de resultados potenciais para causas SUIN.

3 Soluções Práticas para Inferência em Amostras Pequenas

Uma das supostas limitações da pesquisa quali é sua incapacidade de quantificar incerteza. Como vimos, a identificação causal é uma condição do desenho de pesquisa, independente do tamanho amostral. Assim, o próximo passo é, supondo que a pesquisa qualitativa resolveu o problema do desenho de pesquisa, nos perguntar: como fazer inferência em amostras pequenas? A resposta é a estatística Bayesiana, que introduzimos a seguir.

3.1 Inferência Bayesiana

Na estatística Bayesiana, a quantificação da incerteza depende de uma distribuição de probabilidade a priori, que usualmente é interpretada como a incerteza da pesquisadora antes de observar evidência empírica, e uma verossimilhança, que é uma modelagem do processo de geração de dados. O teorema de Bayes, que dá nome à abordagem estatística, nos diz como combinar de maneira rigorosa a priori e a verossimilhança, para obtermos uma distribuição de probabilidade a posteriori (isto é, após observar a evidência empírica) sobre as quantidades de interesse. Não há diferença formal entre n pequeno ou grande. O impacto do tamanho amostral é apenas em reduzir a incerteza a posteriori. É possível mostrar que, sob condições bastante gerais de regularidade, a distribuição a posteriori converge para a distribuição verdadeira à medida que o tamanho amostral vai para infinito. Em outras palavras, a incerteza se reduz com n maior, mas a rigor nada impede de se estimar qualquer quantidade de interesse com um n pequeno.

Na verdade, a questão crítica para a inferência estatística não é simplesmente o tamanho da amostra (n), mas também a relação sinal-ruído; ou seja, quanto mais forte o efeito (o “sinal”) em relação ao ruído (variabilidade dos dados), menos dados são necessários para detectá-lo.

Não é por outra razão que estudos qualitativos bem desenhados geralmente se concentram em contextos onde o sinal é forte e claramente observável, permitindo inferências robustas mesmo com menos casos. Esse tipo de estratégia visa resolver problemas de inferência, mas nada dizem sobre a validade interna, que estão relacionadas às condições formais de identificação.

Por outro lado, estudos quantitativos com grandes amostras podem ser necessários para detectar efeitos menores, onde o sinal é fraco em relação ao ruído. Portanto, a crítica ao tamanho da amostra na verdade é imprecisa, pois é preciso combinar o tamanho amostral com a razão sinal-ruído, para só então avaliar qual tipo de desenho de pesquisa deve produzir menor incerteza inferencial. Sem analisar o caso concreto, nada pode ser dito sobre um tipo de pesquisa ou outro.

E a abordagem Bayesiana é uma das formas pela qual se pode quantificar essa relação sinal-ruído e tamanho amostral. Se o sinal for muito fraco para um dado tamanho amostral, haverá reflexo na quantificação da incerteza que será muito elevada. Se por outro lado o sinal for forte o suficiente para o dado número de casos, a incerteza estimada permitirá inferências e conclusões mais assertivas.

3.1.1 Fundamentos da Probabilidade Bayesiana

Para facilitar a compreensão das novas metodologias que integram inferência Bayesiana com inferência causal em estudos qualitativos, é útil apresentar uma rápida introdução à maquinaria Bayesiana.

3.1.1.1 Teorema de Bayes O Teorema de Bayes permite combinar conhecimento a priori com dados observados para atualizar a crença sobre um parâmetro ou evento. Formalmente, seja H uma hipótese causal, e E a evidência ou conjunto de evidências coletadas. O objetivo é quantificar a probabilidade da hipótese ser verdadeira, dada a evidência.

O teorema de Bayes diz que: $p(H|E) = \frac{p(H)p(E|H)}{p(E)}$. Aqui, $p(H)$ é a probabilidade a priori da hipótese, $p(E|H)$ é a verossimilhança, isto é, uma quantificação de quanto a evidência é mais ou menos provável, dada a hipótese sobre o mecanismo causal, e $p(E)$ é chamada de constante normalizadora e mede a probabilidade a priori da evidência. E $p(H|E)$ é a probabilidade a posteriori da hipótese, condicional as evidências, ou seja, quão provável é uma hipótese diante das evidências.

3.1.1.2 Função de Verossimilhança A função de verossimilhança é a probabilidade dos dados observados tratada como uma função dos parâmetros. Pode ser escrita como

$L(\theta|x_1, x_2, \dots, x_n) = p(x_1, x_2, \dots, x_n|\theta)$, onde os dados são fixos e o parâmetro θ é aleatório. No caso mais simples de uma hipótese, escrevemos $P(E|H)$. O importante é notar que, apesar de escrever P , não se trata propriamente de uma probabilidade, mas de uma verossimilhança.

4 Novos Desenhos Causais Qualitativos

Para que a inferência Bayesiana possa ser útil, é preciso que o desenho de pesquisa seja crível para responder a perguntas causais. Nesta seção, apresentaremos duas soluções recentes que utilizam a metodologia Bayesiana para estimar quantidades causais em pesquisas qualitativas: uma desenvolvida no livro *Integrated Inferences* de Humphreys and Jacobs (2023) (ver também Humphreys and Jacobs (2015)) e outra no livro *Social Inquiry and Bayesian Inference* de Fairfield and Charman (2022) (ver também Fairfield and Charman (2019); Fairfield and Charman (2017)).

Essas soluções deixam claro e sem sombra de dúvidas o argumento do trabalho de que o problema de identificação é completamente distinto e ortogonal ao problema da inferência estatística.

4.1 Process Tracing Bayesiano

O rastreamento de processos bayesiano é uma abordagem robusta para inferência causal em estudos qualitativos. Utilizando a probabilidade bayesiana, podemos atualizar nossas crenças sobre hipóteses causais à medida que novas evidências são coletadas. Este método é particularmente útil em contextos onde as evidências são escassas ou os processos causais são complexos.

Assim, envolve a especificação de hipóteses H_i e suas probabilidades a priori $P(H_i)$. Deve-se identificar as evidências disponíveis e construir a verossimilhança $P(E|H_i)$ para cada hipótese. As comparações de hipóteses são feitas usando as *odds* a posteriori, de forma a não precisa computar a constante normalizadora:

$$\frac{P(H_i|E)}{P(H_j|E)} = \frac{P(H_i)P(E|H_i)}{P(H_j)P(E|H_j)}$$

A seguir explico como cada um dos componentes da fórmula acima (priori e verossimilhança) são formulados na prática.

4.1.1 Definição de Prioris

A definição de *prioris* pode ser feita por meio da elicitacão de experts, uso de *prioris* não-informativas e *prioris* informativas definidas pela pesquisadoras.

A elicitacão de experts envolve a consulta a especialistas e utilização de métodos para transformar o conhecimento deles em distribuições de probabilidade. Como tais pessoas não costumam saber traduzir seus conhecimentos em distribuição de probabilidade, a literatura dessa área tem desenvolvido métodos para extrair esse conhecimento e quantificá-lo em distribuições de probabilidade (O’Hagan (2019), Albert et al. (2012), O’Hagan et al. (2006), O’Hagan (1998)).

Prioris não-informativas são a forma mais básica e tradicional de definição de *prioris* (O’Hagan 2010). A ideia é escolher uma *priori* que torne todas as hipóteses a *priori* igualmente prováveis, seguindo o princípio da razão insuficiente (Kass and Wasserman 1996).

Uma terceira possibilidade é o uso de *prioris* informativas, que representem o conhecimento da pesquisadora, que pode advir de conhecimento substantivo ou de resultados de pesquisas prévias como meta-análises ou estudos quantitativos prévios (Gelman, Simpson, and Betancourt 2017; Gelman 2009).

Como argumenta Bouchat (2023), essa parte do livro é muito menos desenvolvida que a da construção da verossimilhança. Há potencial para avançarmos nessa parte, especialmente porque a *priori* pode servir como uma fonte de regularização, de maneira similar ao Double LASSO em machine learning adaptada para inferência causal (Urminsky, Hansen, and Chernozhukov, n.d.). Por ora, o esperado é a utilização de *prioris* não-informativas, para evitar o risco que a pesquisadora utilize *prioris* infladas ou uma espécie de viés de confirmação. Por outro lado, a vantagem é que as *prioris* são transparentes e permitem colocar em termos claros onde há possível discordância. Além disso, é possível (e recomendável) fazer uma análise de robustez com outras *prioris*, para avaliar se e como as conclusões mudam.

4.1.2 Verossimilhanças

A verossimilhança avalia a plausibilidade de observar uma evidência E sob a hipótese H_i . Os autores recomendam adaptar a ideia de decibéis, que utilizamos para classificar o volume de um som, para avaliar o peso de uma evidência. Ou seja, deveríamos usar logaritmos (os decibéis são calculados com logaritmos) para calibrar o peso das evidências:

$$10 \log_{10} \left(\frac{P(E|H_i)}{P(E|H_j)} \right)$$

Para ver a utilidade de se utilizar decibéis para ponderar o peso das evidências, os autores analisam um exemplo introduzido por Bennet (2015), no qual o autor considera um *smoking gun* se $P(E|H_j) = .05$ e $P(E|H_i) = .2$. Aplicando o logaritmo, isso dá cerca de 6db. Segundo Fairfield and Charman (2017), um volume de 6db é saliente, mas estaria longe de ser um *smoking gun*.

É importante destacar que a verossimilhança não pode contaminar a priori. Ou seja, informações que são na verdade evidências que entrariam na verossimilhança não podem ser utilizadas para formar prioris.

4.1.2.1 Evidências Na análise estatística tradicional, é em geral simples determinar quando uma observação termina e outra começa. No caso de estudos qualitativos, em especial rastreo de processo, esse passo é menos trivial.

A definição de evidências é crucial e pode incluir informações de diferentes fontes ou observações correlacionadas. No Bayesianismo, podem ser consideradas evidências de outros casos similares. Evidências devem ser suficientemente distintas para favorecer hipóteses diferentes.

4.1.3 Hipóteses Rivals

A formulação de hipóteses rivais ajuda a reduzir o viés de confirmação (Fairfield and Charman 2019). Matematicamente, duas hipóteses rivais significam que $P(H_i) + P(H_j) = 1$. Se a pesquisa possui uma única hipótese, é fácil avaliar erroneamente se determinada evidência empírica torna a hipótese mais ou menos plausível. Digamos que estamos interessados em

investigar as causas de um acidente de avião¹. Se minha única hipótese é que foi **apenas** falha mecânica, e encontramos evidências de que a caixa preta do avião registrou altitude errada e que o piloto não tinha dormido nas últimas 24 horas. A primeira evidência suporta a hipótese, enquanto a segunda é neutra. Porém, suponha que a outra hipótese (rival) é que houve uma falha mecânica e falha humana em conjunto. Nesse caso, a primeira hipótese parece menos plausível do que a segunda.

Não basta simplesmente formular uma negação lógica, pois ela inclui infinitas hipóteses potencialmente contraditórias entre si. Ou seja, não basta dizer que uma hipótese é que houve apenas falha mecânica e a hipótese rival é que não houve apenas falha mecânica, pois a negação aqui inclui houve falha humana e não houve falha humana.

Uma das limitações da abordagem de rastreio de processo Bayesiana é quando a pesquisadora considera a possibilidade de múltiplas causas, pois pode ser difícil formulá-las como hipóteses rivais. Mas esse é um passo necessário para que a metodologia possa ser empregada.

A ideia de hipóteses rivais é crítica para estabelecer o tipo de pesquisa causal que em geral é possível na metodologia qualitativa. Como vimos, um desafio é garantir que não há variável omitida relevante. Para Slater and Zibblatt (2013), o possível é avaliar a plausibilidade relativa de hipóteses principais, o que condiz com a sugestão de comparar *posterior odds* de Fairfield and Charman (2022).

Os trabalhos clássicos da área, no meu entender, estão fazendo justamente isso. Considere o trabalho de Skocpol (1979) sobre as revoluções. Como ela mesma afirma na introdução, o estudo dos casos Chineses, Russos e Franceses tem a ver com sua frustração com as teorias então vigentes e os mecanismos causais propostos para explicar esses casos. Em outras palavras, teorias alternativas sugerem hipóteses rivais, que ela acredita não são compatíveis com a evidência empírica dos casos. Sua hipótese causal é produzida em contraste com as rivais, e é considerada mais provável que as outras, diante das evidências. Não se trata, portanto, de um desenho metodológico que permita eliminar a possibilidade de variáveis de confusão à explicação proposta, mas de avaliar quais das hipóteses rivais é mais compatível com a evidência empírica dos casos.

¹Agradeço à Elizabeth Balbachevsky pela sugestão do exemplo

O argumento central aqui é que não é um bom desenho de pesquisa análise de casos com apenas uma hipótese a ser testada. Isso porque a hipótese rival é apenas a negação lógica, e inclui toda e qualquer possível relação causal com algum viés de variável omitida. Se esse tipo de abordagem já foi abandonada como recomendável pela revolução da credibilidade, é igualmente problemático em estudos qualitativos. Porém, um desenho de pesquisa que avalie a plausibilidade de hipóteses rivais contorna o problema da variável omitida. Não se trata de mostrar que a hipótese explicativa mais provável não é confundida por alguma variável omitida. Mas mostrar que, diante das principais hipóteses explicativas sugeridas pela literatura, ela é a mais provável. Isso é, em meu entender, exatamente o que Skocpol fez em sua obra clássica. Se houver uma nova hipótese sugerida no futuro, pode ser comparada diante de evidências antigas ou novas e avaliada.

4.2 Inferências Integradas

O trabalho de Humphreys and Jacobs (2023) apresenta o que eles chamam de *queries* causais, que consiste na utilização de Directed Acyclic Graphs (DAGs), tal como desenvolvida por Pearl, juntamente com inferência Bayesiana, para estimar quantidades causais com métodos qualitativos. O termo inferência integrada é porque, ao utilizar metodologias originariamente quantitativas, permite integrar quali-quant, produzindo assim desenhos de pesquisa com métodos mistos. Aqui interessa para nós o aspecto qualitativo da abordagem.

Um primeiro aspecto que a metodologia de Humphreys and Jacobs (2023) deixa claro é que é necessário, para fins práticos, considerar todas as variáveis da pesquisa quali como binárias. Isso faz sentido, posto que seria inviável mensurar quantidades contínuas e relacioná-las estatisticamente com amostras pequenas. Ou seja, a variância potencial seria muito grande, tornando inviável a estimação.

Um segundo aspecto é que, como é costumeiro, é mais fácil modelar as relações causais como determinísticas, ou seja, um caso possui ou não relação causal entre variáveis, não cabendo pensar probabilisticamente. Isso, contudo, não é uma limitação muito relevante, pois a quantificação da incerteza via Bayes significa que poderemos falar de probabilidade do efeito causal existir. Ou seja, a relação é determinística, mas o nosso conhecimento sobre

a relação causal é probabilístico, o que é compatível com boa parte das ontologias sociais.

A suposição de que a relação causal é determinística e binária leva a ideia de que cada caso pode ser classificado em um tipo causal. Os tipos causais podem ser classificados em quatro categorias: adverso, benéfico, crônico e destinado. Cada tipo representa uma relação causal distinta entre tratamento e resposta. Por exemplo, indivíduos adversos melhorariam apenas se não recebessem o tratamento, enquanto os benéficos melhorariam apenas se o recebessem. Os crônicos permanecerão doentes independentemente do tratamento, e os destinados melhorarão independentemente do tratamento. Esse modelo pode ser facilmente generalizado para múltiplas causas, com uma pequena modificação na notação, a qual o leitor interessado pode encontrar no livro de Humphreys e Jacobs.

Para aplicar essa estrutura teórica, utilizamos modelos de queries causais. Esses modelos nos permitem investigar efeitos causais médios, causas ao nível de casos (case-level causal analysis) e caminhos causais. Uma query causal é um algoritmo que recebe um modelo M como entrada (especificado na forma de um DAG) e retorna uma quantidade $Q(M)$ como saída. Existem quatro queries causais chave que são fundamentais para essa análise e respondem a questões distintas: - Efeitos causais ao nível de caso (case-level). - Atribuição causal ao nível de caso. - Efeitos causais médios (ATE). - Caminhos causais.

4.2.1 Efeito Causal ao Nível do Caso

Perguntas de pesquisa frequentemente investigam se X tem um efeito causal em Y em casos específicos. Por exemplo, se houvesse uma crise econômica no Brasil hoje, Lula sofreria impeachment? Em termos contrafactuais, a questão é se, ao intervir exogenamente no valor de X , o valor de Y mudaria. Ou seja, qual seria o tipo causal do Brasil sob Lula III? Se não haveria impeachment com ou sem crise, e considerando impeachment como 1 e não-impeachment como 0, então seria um tipo crônico, pois o impedimento não seria observado com e sem crise econômica. Em outras palavras, poderíamos estar interessados na probabilidade de ser crônico. Ou talvez a pergunta de pesquisa seria em torno da probabilidade do país ser benéfico, isto é, o impeachment ocorreria com crise econômica, mas não em sua ausência. Neste modelo, o estimando corresponde à proporção de casos do tipo crônico em

uma população definida pelo conjunto de casos comparáveis.

Esse tipo de query nos ajuda a entender as implicações de mudanças específicas em variáveis causais. O modelo de queries causais aqui apresentado permite responder a esse tipo de pesquisa, tipicamente qualitativas, de maneira quantitativamente precisa, a partir de estudo de caso.

4.2.2 Atribuição Causal ao Nível do Caso

Além de investigar se X tem um efeito causal, também é importante determinar se X de fato causou Y em casos específicos. Por exemplo, a crise econômica causou o impeachment da Dilma em 2016? Em termos contrafactuais, a questão é se, dado os valores que X e Y assumiram, o valor de Y seria diferente se o valor de X fosse diferente. Este tipo de análise é crucial para entender a relação direta entre causas e efeitos em contextos específicos.

4.2.3 Caminhos Causais

Em casos onde estabelecemos que X causou Y , é relevante investigar se isso ocorreu via um mediador M . Em notação de resultados potenciais, perguntamos se $Y(X = x, M = M(X = 1))$ é maior do que $Y(X = x, M = M(X = 0))$. Este tipo de query nos ajuda a mapear as vias por meio das quais os efeitos causais se manifestam.

4.3 Comparando as Abordagens

As duas metodologias aqui apresentadas são exemplos de desenvolvimento recentes de utilização de inferência Bayesiana para estudos qualitativos. Os trabalhos de Fairfield e Charman têm como objetivo formalizar o rastreamento de processos, metodologia qualitativa que tem se desenvolvido bastante nas últimas décadas. Ela é particularmente útil quando as hipóteses são de fato rivais e permite avaliar hipóteses causais concorrentes para explicar fenômenos sociais nos quais apenas o rastreamento de processo poderá recuperar o efeito causal de interesse.

O modelo de queries causais de Humphreys e Jacobs é mais genérico, podendo ser aplicado a muitos diferentes tipos de pesquisas qualitativas, incluindo rastreamento de processo.

Também utiliza inferência Bayesiana para estimação e quantificação da incerteza, mas não possui a limitação de restringir-se a situações de hipóteses rivais, admitindo hipóteses complementares. Por outro lado, ao não incorporar a ideia de decibéis para ponderar o peso das evidências, perde uma heurística importante do método advogado por Fairfield e Charman. Por fim, com frequência o DAG contempla a possibilidade de variáveis não-observadas que confundem o efeito causal de interesse e, na prática, continuamos com incerteza se o efeito causal é real ou não. Esse, é claro, é o problema apontado na literatura em estudos comparados de que é difícil evitar o viés de variável omitida em estudos comparados controlados. A solução de Fairfield and Charman (2022) de hipóteses rivais é mais promissora.

Deve ficar claro, contudo, que estudos qualitativos baseados na comparação de hipóteses rivais não necessariamente garantem que o efeito causal é sem viés, pelos argumentos já reconhecidos na literatura sobre a possibilidade de variável omitida. Por outro lado, ao induzir esparsividade de hipóteses causais, como no método de Double Lasso, é possível pelo menos avaliar como as evidências permitem descartar hipóteses teóricas menos consistentes com as evidências disponíveis, na linha da *Inference to the Best Explanation* (Spirling and Stewart 2025).

Como são metodologias novas, aperfeiçoamentos e mesmo a mistura de ambas podem vir a ser desenvolvimentos possíveis. Por hora, importa salientar que ambas são promissoras e o método de process tracing Bayesiano já está sendo aplicado (Fairfield and Charman 2025; Rabbia 2023), inclusive com alunos que viram o conteúdo em aula e já disseram que estão usando em seus trabalhos (ainda não publicados). Contudo, é esperado que seja necessário treinamento mais longo e extenso, como os próprios autores reconhecem em seus trabalhos.

5 Transportabilidade, Generalização ou Validade Externa

A preocupação com identificação causal, destacada pela revolução da credibilidade, implicou uma ênfase na validade interna dos estudos. A validade interna se refere à capacidade de um estimador identificar um efeito causal (estimando) sem viés na sua amostra específica. Por

outro lado, a validade externa refere-se à capacidade de generalizar os achados para outras populações e contextos (Findley, Kikuta, and Denly 2021; Muller 2015)]. A literatura qualitativa metodológica tem feito confusão sobre os conceitos de validade interna, misturando generalização de uma amostra para a população como validade externa Slater and Ziblatt (2013).

Potanto, de saída podemos descartar como infundadas duas posições. Se, como a literatura metodológica argumenta, pesquisa qualitativa não tem maiores problemas em atingir validade interna, então não há qualquer diferença com relação a esse aspecto em métodos qualitativos ou quantitativos.

Similarmente, as duas metodologias são igualmente limitadas (a princípio) em sua capacidade de possuírem validade externa. E essas duas conclusões seguem do fato de que não há relação entre validade interna e externa e tamanho amostral, se a pesquisa é em profundidade ou não, ou quaisquer das possíveis distinções empregadas para diferenciar métodos qualitativo de quantitativo.

Como discutido anteriormente, a inferência causal depende inicialmente da satisfação das condições para identificação causal. Uma vez resolvida essa questão, a inferência estatística pode ser feita por meio de métodos bayesianos apropriados para estudos com pequeno n . Contudo, isso garante apenas a validade interna do estudo, ou seja, não permite, por si só, generalizar os resultados para outros contextos ou populações, o que caracteriza a validade externa.

Nesse sentido, a suposta incapacidade de generalização devido ao pequeno número de casos analisados não tem fundamento sólido. Tanto estudos quantitativos quanto qualitativos geralmente compartilham as mesmas limitações quanto à generalização, refletidas na ausência de garantias explícitas sobre sua validade externa.

A distinção entre validade interna e externa foi introduzida por Campbell (1957) no contexto de estudos experimentais. Na definição original, validade interna referia-se à capacidade de um experimento detectar o efeito causal de interesse na amostra estudada. Atualmente, consideramos que um estudo (não apenas experimental) tem validade interna quando seu desenho, execução (incluindo coleta de dados) e análise permitem responder adequadamente

às perguntas de pesquisa para a amostra analisada (McDermott (2011)).

A validade externa por outro lado, diz respeito a transportar os efeitos estimados de uma população a outra Pearl and Bareinboim (2022). No acabouço de resultados potenciais, a validade externa pode ser definida do seguinte modo. Seja D_i uma variável binária que indica se uma unidade está na amostra ($D_i = 0$) ou na população de interesse ($D_i = 1$). Dizemos que existe validade externa se $\mathbb{E}[Y_i(1) - Y_i(0)|D_i = 0] = \mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1]$. Em palavras, a diferença média nos resultados potenciais do tratamento e controle na população de interesse é igual ao da população de onde foi retirada a amostra. Com a utilização de DAGs, a literatura tem mostrado que o essencial é especificar quais variáveis da população alvo podem modificar o efeito causal em relação à população estudada.

No simpósio da QMMR de 2023 sobre o livro de Fairfield e Charman (Fairfield and Charman 2022), Jacobs (A. M. Jacobs 2023) e Fairfield e Charman (Fairfield and Charman 2023) discutem a questão da transportabilidade (sob o nome de generalização) a partir de um caso presente no livro de Fairfield and Charman (2022): qual a causal da mobilização democrática no sudeste asiático autoritário? Três hipóteses são consideradas: 1. elites comunais autônomas promovendo mobilização; 2. declínio econômico como o fator central e 3. eleições fraudadas. Ao analisar evidência de Filipinas e Vietnã, concluem que a hipótese mais provável é a de elites comunicais. Como Jacobs observa, dada a condição de escopo, a plausibilidade dessa explicação se aplica igualmente a todos os países autoritários do sudeste asiático como Tailândia e Malásia autocráticas.

A resposta de (Fairfield and Charman 2023), de que podemos contrapor outra hipótese com outra condição de escopo, parece-nos insuficiente. No exemplo dado pelos autores, se a hipótese rival for “The communal-elites causal logic operates in the Philippines and Vietnam, whereas economic decline instead sparks democratic mobilization elsewhere in Southeast Asia”, evidência da Filipinas apenas será incapaz de informar ou diferenciar entre as hipóteses.

A recomendação de Fairfield e Charman sobre generalização, que consiste em redefinir continuamente as condições de escopo ao se contrapor novas hipóteses, apresenta problemas fundamentais para a generalização rigorosa. A lógica de ajustar o escopo a cada nova evidência torna virtualmente impossível estabelecer conclusões generalizáveis estáveis, pois sempre

será possível introduzir uma hipótese alternativa restrita a um novo escopo ainda não testado. Por exemplo, suponha que após a análise de Filipinas e Vietnã, evidências adicionais sejam coletadas na Malásia, reforçando a hipótese original de elites comunais. Mesmo com essa nova evidência, é possível criar imediatamente uma hipótese ad hoc afirmando que as elites comunais explicam Filipinas, Vietnã e Malásia, mas que outro fator, como declínio econômico, seria decisivo em outros países ainda não analisados.

Ademais, a definição de escopo pode incluir elementos arbitrários como períodos específicos, níveis de desenvolvimento econômico de países vizinhos ou outros fatores contextuais ilimitados. Como é impraticável em pesquisas qualitativas levar em consideração todas as possíveis variáveis contextuais, isso implica, na prática, que nenhuma generalização robusta seria de fato alcançável sob a abordagem de escopo ajustável proposta pelos autores.

Portanto, a recomendação de Fairfield e Charman, embora intuitivamente atraente por sua flexibilidade, resulta em um critério excessivamente permissivo. Isso impede a generalização real dos resultados qualitativos, pois as conclusões ficam continuamente suscetíveis à redefinição arbitrária do escopo em resposta a qualquer evidência contraditória futura. Em contraste, a perspectiva aqui defendida sugere a necessidade de teorizar explicitamente sobre os mecanismos causais que definem os limites do escopo de aplicação, fornecendo critérios claros e testáveis para avaliar a validade da generalização.

No exemplo acima, a pesquisa qualitativa que pretende generalizar para além dos casos estudados deve teorizar sobre que mecanismo faria Filipinas e Vietnã serem diferentes do restante do sudeste asiático. Em outras palavras, não é uma boa prática simplesmente afirmar hipóteses distintas modificando o escopo, mas é necessário teorizar o que determina que um escopo seja válido ou não. Retornando ao exemplo do trabalho de Skocpol, suas condições de escopo não são apenas os casos de China, Rússia e França, mas o caso de países agrários, não-colonizados recentemente e com estados proto-burocráticos enfrentando adversários economicamente desenvolvidos. Isso significa que, se formos transportar essa explicação para, digamos, Cuba, teremos que avaliar em que medida um estado parcialmente colonizado e pouco burocrático terá condições diferentes e explicações potencialmente diferentes.

Como esse exemplo mostra, a generalização é difícil de ser realizada e as soluções propostas

na literatura estão longe de serem ponto pacífico. Contudo, tampouco é algo resolvido na abordagem influenciada pela revolução da credibilidade. Se algo, por hora parece mais fácil endereçar essa questão na pesquisa qualitativa do que na quantitativa, como mostramos acima. Mas é a prática aplicada que irá dizer em que medida cada abordagem irá lidar com esse problema.

6 Considerações Finais

A pós o livro de KKV, consolidou-se a percepção (falsa) de que métodos qualitativos seriam mais limitados em sua capacidade de generalização, embora superiores em sua validade interna. Essa percepção, baseada na adoção da perspectiva então vigente de modelos de regressão quantitativos, dependia da ideia errônea de que inferência causal e estatística eram a mesma coisa.

A Revolução da Credibilidade, contudo, trouxe uma nova perspectiva, segundo a qual a identificação causal (garantir a validade interna de um estudo) é completamente ortogonal à inferência estatística. Como resultado, mostramos que não há nenhuma característica intrínseca à pesquisa quali ou quanti que justifique diferenças quanto à validade interna e externa das respectivas pesquisas.

Desfeita essa confusão, mostramos que a inferência Bayesiana oferece soluções robustas para a inferência em amostras pequenas, equiparando o rigor das pesquisas qualitativas ao das quantitativas. As ilustrações metodológicas, como o *process tracing* Bayesiano e as *queries* causais, exemplificam como essas novas abordagens podem ser aplicadas na prática. Esses métodos permitem uma análise mais detalhada e precisa, reforçando a validade dos estudos qualitativos.

Por fim, abordamos como a literatura baseada na ideia de transportabilidade pode permitir que a pesquisa qualitativa supere suas limitações quanto à validade externa.

Assim, o trabalho contribui para tornar mais claras as dificuldades e caminhos da pesquisa qualitativa baseada em casos, desfazendo qualquer pretensão de superioridade intrínseca da pesquisa quantitativa causal. Esperamos que este estudo inspire uma reavaliação das aborda-

gens metodológicas nas ciências sociais, promovendo um debate mais equilibrado e integrador.

7 References

- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2015. "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science* 59 (2). Wiley Online Library: 495–510.
- Albert, Isabelle, Sophie Donnet, Chantal Guihenneuc-Jouyaux, Samantha Low-Choy, Kerrie Mengersen, and Judith Rousseau. 2012. "Combining Expert Opinions in Prior Elicitation."
- Amorim, Octavio, and Júlio César Cossio Rodriguez. 2016. "O Novo método Histórico-Comparativo e Seus Aportes à Ciência Política e à Administração pública." *Revista de Administração Pública* 50 (6). SciELO Brasil: 1003–27.
- Angrist, Joshua D., and Alan B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106 (4). MIT Press: 979–1014.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton university press.
- . 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics." *Journal of Economic Perspectives* 24 (2). American Economic Association: 3–30.
- Bennett, Andrew, and Jeffrey T. Checkel. 2015. *Process Tracing*. Cambridge University Press.
- Blatter, Joachim, and Markus Haverland. 2012. *Designing Case Studies: Explanatory Approaches in Small-N Research*. Springer.
- Bouchat, Sirus. 2023. "Social Inquiry and Bayesian Inference: An "Objective" Vision for Mixed Method Research?" November. Zenodo. doi:10.5281/ZENODO.8326432.
- Brady, Henry E, and David Collier. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Rowman & Littlefield Publishers.
- Brady, Henry E, David Collier, and Jason Seawright. 2010. "Refocusing the Discussion of Methodology." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, 15–31. Rowman and Littlefield.
- Campbell, Donald T. 1957. "Factors Relevant to the Validity of Experiments in Social Settings." *Psychological Bulletin* 54 (4).
- Card, David. 2022. "Design-Based Research in Empirical Microeconomics." *American Economic Review* 112 (6). American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203: 1773–81.
- Card, David, and Alan B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *The American Economic Review* 84 (4): 772–93.
- Collier, David. 2011. "Understanding Process Tracing." *PS: Political Science & Politics* 44 (4). Cambridge University Press: 823–30.
- Collins, Kathleen M. T. 2015. *Validity in Multimethod and Mixed Research*. Edited by Sharlene Nagy Hesse-Biber and R. Burke Johnson. Oxford University Press. doi:10.1093/oxfordhb/9780199933624.013.17.
- Fairfield, Tasha, and Andrew Charman. 2019. "A Dialogue with the Data: The Bayesian Foundations of Iterative Research in Qualitative Social Science." *Perspectives on Politics* 17 (1). Cambridge University Press: 154–67.
- . 2022. *Social Inquiry and Bayesian Inference*.

- . 2023. “Bayesian Reflections,” November. Zenodo. doi:10.5281/ZENODO.8326443.
- . 2025. “Bayesian Reasoning for Qualitative Replication Analysis: Examples from Climate Politics.” *Political Science Research and Methods*. Cambridge University Press, 1–16.
- Fairfield, Tasha, and Andrew E. Charman. 2017. “Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities, and Caveats.” *Political Analysis* 25 (3). Cambridge University Press: 363–80.
- Findley, Michael G., Kyosuke Kikuta, and Michael Denly. 2021. “External Validity.” *Annual Review of Political Science* 24 (1): 365–93. doi:10.1146/annurev-polisci-041719-102556.
- Forozish, Ali Omar. 2024. “How the Credibility Revolution Created a Paradigm Shift.” *Available at SSRN 4744474*.
- Gelman, Andrew. 2009. “Bayes, Jeffreys, Prior Distributions and the Philosophy of Statistics.” *Statistical Science* 24 (2). JSTOR: 176–78. <https://www.jstor.org/stable/25681293>.
- Gelman, Andrew, Daniel Simpson, and Michael Betancourt. 2017. “The Prior Can Often Only Be Understood in the Context of the Likelihood.” *Entropy* 19 (10). MDPI: 555.
- George, Alexander L, and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. mit Press.
- Glied, Sherry. 2021. “The Credibility Revolution in Economics and How It Has Changed Health Policy.” In *JAMA Health Forum*, 2:e214335–35. American Medical Association.
- Goldsmith-Pinkham, Paul. 2024. “Tracking the Credibility Revolution Across Fields.” arXiv. <https://arxiv.org/abs/2405.20604>.
- Grimmer, Justin. 2015. “We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together.” *PS: Political Science & Politics* 48 (1). Cambridge University Press: 80–83.
- Humphreys, Macartan, and Alan M. Jacobs. 2015. “Mixing Methods: A Bayesian Approach.” *American Political Science Review* 109 (4). Cambridge University Press: 653–73.
- . 2023. “Integrated Inferences.” *Null*. doi:10.1017/9781316718636.
- Jacobs, Alan. 2022. “Counterfactuals, Mechanisms, and Background Beliefs in The Logic of Social Science,” April. Zenodo. doi:10.5281/ZENODO.6448178.
- Jacobs, Alan M. 2023. “Leaning in to Analytic Explicitness,” November. Zenodo. doi:10.5281/ZENODO.8326449.
- Kass, Robert E., and Larry Wasserman. 1996. “The Selection of Prior Distributions by Formal Rules.” *Journal of the American Statistical Association* 91 (435): 1343–70. doi:10.1080/01621459.1996.10477003.
- Keele, Luke. 2015a. “The Discipline of Identification.” *PS: Political Science & Politics* 48 (1). Cambridge University Press: 102–6.
- . 2015b. “The Statistics of Causal Inference: A View from Political Methodology.” *Political Analysis* 23 (3). Cambridge University Press: 313–35.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press. doi:10.1515/9781400821211.
- Leamer, Edward E. 1983. “Let’s Take the Con Out of Econometrics.” *The American Economic Review* 73 (1). JSTOR: 31–43. <https://www.jstor.org/stable/1803924>.
- Libman, Alexander. 2023. “Credibility Revolution and the Future of Russian Studies.” *Post-Soviet Affairs* 39 (1-2): 60–69. doi:10.1080/1060586X.2022.2148446.

- Lundberg, Ian, Rebecca Johnson, and Brandon M. Stewart. 2021. "What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory." *American Sociological Review* 86 (3): 532–65. doi:10.1177/00031224211004187.
- Mahoney, James. 2008. "Toward a Unified Theory of Causality." *Comparative Political Studies* 41 (4-5). Sage Publications Sage CA: Los Angeles, CA: 412–36.
- Mahoney, James, and Gary Goertz. 2006. "A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research." *Political Analysis* 14 (3): 227–49. doi:10.1093/pan/mpj017.
- McDermott, Rose. 2011. "Internal and External Validity." *Cambridge Handbook of Experimental Political Science* 27.
- Muller, Seán M. 2015. "Causal Interaction and External Validity: Obstacles to the Policy Relevance of Randomized Evaluations." *The World Bank Economic Review* 29 (suppl 1): S217–25. doi:10.1093/wber/lhv027.
- O'Hagan, Anthony. 1998. "Eliciting Expert Beliefs in Substantial Practical Applications." *Journal of the Royal Statistical Society Series D: The Statistician* 47 (1). Oxford University Press: 21–35.
- . 2010. *Kendall's Advanced Theory of Statistic 2B*. John Wiley & Sons.
- . 2019. "Expert Knowledge Elicitation: Subjective but Scientific." *The American Statistician* 73 (sup1): 69–81. doi:10.1080/00031305.2018.1518265.
- O'Hagan, Anthony, Caitlin E. Buck, Alireza Daneshkhah, J. Richard Eiser, Paul H. Garthwaite, David J. Jenkinson, Jeremy E. Oakley, and Tim Rakow. 2006. "Uncertain Judgements: Eliciting Experts' Probabilities." John Wiley & Sons.
- Pearl, Judea, and Elias Bareinboim. 2011. "Transportability of Causal and Statistical Relations: A Formal Approach." In *Proceedings of the AAAI Conference on Artificial Intelligence*, 25:247–54.
- . 2022. "External Validity: From *Do* -Calculus to Transportability Across Populations." In *Probabilistic and Causal Inference*, edited by Hector Geffner, Rina Dechter, and Joseph Y. Halpern, 1st ed., 451–82. New York, NY, USA: ACM. doi:10.1145/3501714.3501741.
- Rabbia, Martin. 2023. "Why Did Argentina and Uruguay Decide to Pursue a Carbon Tax? Fiscal Reforms and Explicit Carbon Prices." *Review of Policy Research* 40 (2). Wiley Online Library: 230–59.
- Ragin, Charles C. 2014. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Univ of California Press.
- Rihoux, Benoît, and Charles C. Ragin. 2009. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Vol. 51. Sage.
- Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies." *Journal of Educational Psychology* 66 (5). American Psychological Association: 688.
- Samii, Cyrus. 2016. "Causal Empiricism in Quantitative Research." *The Journal of Politics* 78 (3): 941–55. doi:10.1086/686690.
- Seawright, Jason. 2018. "Beyond Mill: Why Cross-Case Qualitative Causal Inference Is Weak, and Why We Should Still Compare," March. Zenodo. doi:10.5281/ZENODO.2562153.
- Sekhon, Jasjeet S. 2004. "Quality Meets Quantity: Case Studies, Conditional Probability, and Counterfactuals." *Perspectives on Politics* 2 (02). doi:10.1017/S1537592704040150.
- Simmons, Erica S., Nicholas Rush Smith, and Rachel A. Schwartz. 2018. "Rethinking

- Comparison in the Social Sciences,” March. Zenodo. doi:10.5281/ZENODO.2562143.
- Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia and China*. Cambridge University Press.
- Slater, Dan, and Erica Simmons. 2010. “Informative Regress: Critical Antecedents in Comparative Politics.” *Comparative Political Studies* 43 (7): 886–917. doi:10.1177/0010414010361343.
- Slater, Dan, and Daniel Ziblatt. 2013. “The Enduring Indispensability of the Controlled Comparison.” *Comparative Political Studies* 46 (10). SAGE Publications Sage CA: Los Angeles, CA: 1301–27.
- Soifer, Hillel David. 2019. “Units of Analysis in Subnational Research.” *Inside Countries: Subnational Research in Comparative Politics* 92. Cambridge University Press Nueva York.
- Spirling, Arthur, and Brandon M. Stewart. 2025. “What Good Is a Regression? Inference to the Best Explanation and the Practice of Political Science Research.” *The Journal of Politics*, June, 000–000. doi:10.1086/734280.
- Urminsky, Oleg, Christian Hansen, and Victor Chernozhukov. n.d. “The Double-Lasso Method for Principled Variable Selection.” OSF.