



12º Encontro da ABCP

19 a 23 de Outubro de 2020

Evento Online

Área Temática: Ensino e Pesquisa em Ciência Política e Relações Internacionais

**MODELOS DE INFERÊNCIA CAUSAL: STRUCTURAL CAUSAL MODELS E POTENTIAL  
OUTCOMES EM PERSPECTIVA COMPARADA**

Matheus Cunha

Universidade Federal de Pernambuco

Caio Rios

Universidade Federal de Pernambuco

Manoel Galdino

Transparência Brasil

Davi Moreira

Faculdade Getúlio Vargas

## Resumo

Como se sabe que dois fenômenos possuem uma conexão causal? A Ciência discute a questão há séculos. Atualmente, a forma mais sofisticada de lidar com o problema é através da construção de modelos causais, que desenvolvem linguagem matemática para representar o que se chama de problema da identificação. Neste artigo, apresentamos dois modelos de inferência causal: *Potential Outcomes*, que é dominante nas Ciências Sociais, e *Structural Causal Models*, cuja concepção moderna advém da Ciência da Informação e Epidemiologia. Ambos apresentam noções contrafactuais de causalidade. Discutimos as concepções centrais para cada um desses modelos e (1) como eles lidam com as ameaças mais prevalentes à identificação, incluindo *confounding* e causalidade reversa, e (2) como eles auxiliam na produção de desenhos de pesquisa sólidos. Nosso argumento central é que existe uma série de pontos de complementaridade entre os dois modelos, o que sugere que sua utilização conjunta é ainda mais benéfica à produção científica. Discutimos esses pontos e ilustramos nosso argumento com a análise de dois artigos empíricos dentre os mais citados em periódicos de Ciência Política na América Latina.

Palavras-Chave: Causação; Potential Outcomes; Structural Causal Models; Directed Acyclic Graphs

## Abstract

How can we know if two phenomena share a causal connection? Science has been discussing the question for centuries. Currently, the most sophisticated way of dealing with the issue is through the elaboration of causal models, which develop mathematical language to represent what is called the identification problem. In this paper, we present two models of causal inference: Potential Outcomes, which is dominant in the Social Sciences, and Structural Causal Models, whose latest conception originated in Computer Science and Epidemiology. Both models present counterfactual notions of causality. We discuss central conceptions of each model and (1) how they deal with threats to identification, such as confounding and reverse causation, and (2) how they aid in producing sound research designs. Our main argument is that there is a series of complementarities between both models, which suggest that their joint implementation is even more beneficial for scientific research. We discuss these complementarities and illustrate our argument with the analysis of two empirical papers among the most cited in Political Science journals in Latin America.

Palavras-Chave: Causation; Potential Outcomes; Structural Causal Models; Directed Acyclic Graphs

## Introdução

Algumas causas são óbvias e facilmente observáveis após alguma manipulação do ambiente. Por exemplo, a força aplicada sobre um objeto explica a distância por ele percorrida. Ao observar um grupo de pessoas arremessando pedras do mesmo lugar, é possível concluir que o mais forte arremessará mais longe<sup>1</sup>. Algumas causas óbvias são falsas. No século XIX, em Viena, Ignaz Semmelweis foi defrontado com um problema médico comum na época: a febre puerperal. Em particular, essa febre acometia mulheres que tinham acabado de dar à luz. No Hospital Geral de Viena, as grávidas podiam parir com a assistência de parteiras ou com os médicos do hospital. Semmelweis observou que a taxa de mortalidade na clínica dos médicos era desproporcionalmente maior do que na sala ocupada pelas parteiras. O curioso é que essa discrepância tinha uma explicação óbvia: a febre puerperal era considerada um tipo de miasma, de acordo com teorias derivadas dos trabalhos de Hipócrates. Condições atmosféricas, fatores climáticos e o contágio produzido pela disseminação dos miasmas de uma mulher grávida para outras provocavam altas taxas de contaminação e decorrente mortalidade. Além disso, alguns médicos acreditavam que o tratamento dado às grávidas era muito severo, o que as tornavam suscetíveis ao miasma (Tyler-Smith, 1856; Wyclicky e Skopec, 1983).

Semmelweis eventualmente associou a ocorrência da febre aos hábitos higiênicos dos médicos, que na melhor das hipóteses lavavam as mãos apenas com água. Muitos trafegavam entre áreas do hospital, lidando com infecções, ferimentos e até cadáveres, posteriormente realizando partos com as mãos nuas. As bactérias acumuladas nas mãos encontravam um ambiente propício no canal vaginal, deflagrando a infecção. Em 1840, a Bacteriologia ainda não havia se tornado uma concepção dominante; trabalhos com microscópios eram raros. Semmelweis morreu aos 47 anos em um asilo mental, tendo sua teoria como alvo de críticas e desacreditada.

Identificar causas não é trivial. Neste artigo, definimos uma causa como **um fenômeno, estrutura, ação ou força que provoca uma mudança em outras estruturas, ações, fenômenos e forças**. Adicionalmente, uma causa **pode ocasionar o surgimento desses elementos**. A partir dessa definição e exemplos apresentados, percebe-se que o problema central de causalidade é conectar uma suposta causa a um efeito observado. Esse tem sido o

---

<sup>1</sup> O exemplo é meramente ilustrativo. Noções de causalidade em algumas disciplinas das Ciências Naturais possuem diferenças fundamentais com as Ciências Sociais, Ciências Médicas e demais disciplinas, dada a natureza dos objetos estudados e teorias construídas.

esforço tanto de cientistas quanto de filósofos da Ciência há séculos. Hume afirmou que não era possível estabelecer relações de causa e efeito, pois somente regularidades empíricas podem ser observadas. Este argumento ficou conhecido também como o problema da indução, isto é, como generalizar relações de causa e efeito apenas a partir de evidência empírica. Outros apontaram para a necessidade de sequenciamento temporal: causas sempre antecedem efeitos (Hume, 1739). As Ciências Naturais do século XIX beberam dessas ideias, incorporando-as à busca por leis determinísticas.

É possível argumentar também que as Ciências Sociais nasceram com um propósito causal. Estudiosos como Durkheim tinham uma veia positivista, utilizando correlações e dados quantitativos para evidenciar a presença de leis estruturais. Essa corrente inicial não demorou a ser contestada por anti-positivistas e demais correntes sociológicas, as quais argumentavam que o mundo social diferia fundamentalmente do mundo físico (Durkheim, 1895; Ashley e Orestein, 2005). Enquanto alguns sociólogos, a exemplo de Weber, ainda defendiam que a Sociologia poderia ser chamada de ciência, devido à capacidade de observar associações entre fenômenos, a disciplina adquiriu um forte componente interpretativo, bem evidenciado no interacionismo simbólico, construtivismo social e teoria crítica.

De forma mais geral, a Ciência sofreu uma série de transformações, especialmente a partir do século XX, sobre o que constituía uma causa, quais abordagens poderiam ser classificadas como científicas e como atestar a validade de uma metodologia científica específica. Em particular, a abordagem contrafactual floresceu como perspectiva dominante de compreensão dos problemas elementares da identificação causal. Nas Ciências Sociais (e em uma série de outras disciplinas), essa perspectiva foi codificada no framework conhecido como Potential Outcomes, representado pelo modelo Neyman-Rubin-Holland (NRH) de agora em diante) de inferência causal (Neyman, 1923; Rubin, 1974; Holland, 1986; Brady, 2008). Esse modelo é altamente influente, e serviu como instrumento para a revolução de credibilidade na Economia (Angrist e Pischke, 2010) e potencialmente na Ciência Política também a partir de 1994. Neste artigo, apresentamos a lógica proposta pelo modelo, seu potencial explicativo e como ele modela problemas habituais relativos à inferência.

O modelo NRH não foi adotado em todas as áreas do conhecimento. A Epidemiologia e a Ciência da Informação, por exemplo, utilizam uma abordagem até pouco tempo largamente ausente em discussões metodológicas nas Ciências Sociais. Essa abordagem pode ser classificada como abordagem de DAGs, porém é mais amplamente conhecida como Structural

(SCM de agora em diante) mudou o estilo: apresentar a abreviatura no texto ou em rodapé.

Causal Models<sup>2</sup>. Judea Pearl (2009), dentre outros, desenvolveu a ideia de gráficos causais para criar um framework causal inteiro. SCM também partem de uma perspectiva contrafactual de causalidade, mas empregam uma linguagem causal distinta, assim como uma codificação própria de desenhos de pesquisa. Mais recentemente, a publicação do The Book of Why (Pearl e Mackenzie, 2018) serviu como introdução desse framework causal nas Ciências Sociais. Em geral, cientistas sociais parecem céticos em relação aos benefícios (Imbens, 2019). Algumas razões para a rejeição são o fato de que nada na linguagem causal de DAGs não pode ser representado via a notação padrão de Potential Outcomes e que desenhos de pesquisa úteis para as Ciências Sociais foram ~~todos~~ desenvolvidos pelo modelo NRH. referentes a todos os tipos de...

Neste artigo, argumentamos: (1) que essas críticas são fundamentalmente inválidas<sup>3</sup>, e (2) que Potential Outcomes e SCM são melhor compreendidos como frameworks complementares. Questões elementares de inferência podem ser apresentadas nos dois modelos, contudo SCM oferecem uma facilidade maior na codificação de pressupostos de pesquisa e representação dos desenhos propostos. Por sua vez, a ênfase dada ao problema da comparabilidade entre grupos no modelo NRH traz atenção a quais diferenças específicas produzem os estimadores dos efeitos.. Descrevemos pontos de complementaridade no artigo, ilustrando-os com dois exemplos concretos em que aplicamos ambas lógicas em relação a dois artigos de alta qualidade publicados por cientistas sociais. Como também demonstramos, neste artigo, preocupações de causalidade ainda são flagrantemente ausentes na pesquisa social brasileira. Esperamos (re)introduzir frameworks completos e úteis para a produção de pesquisa empírica, além de encorajar pesquisadores a pensar mais sistematicamente acerca de seus objetos de pesquisa.

O restante deste artigo está estruturado da seguinte forma: (1) uma explicação acerca de Potential Outcomes, (2) uma explicação sobre Structural Causal Models, (3) o argumento central, (4) a apresentação de dados sobre pesquisa social aplicada no Brasil, e (5) a ilustração do argumento nos dois artigos selecionados.

## Modelos de Causalidade

---

<sup>2</sup> Durante o artigo, utilizamos *Structural Causal Models* (SCM) e *Directed Acyclic Graphs* (DAGs) intercambivelmente.

<sup>3</sup> Por exemplo, a complexidade de um modelo causal não é proibitiva em linguagem de DAGs, mas torna-se rapidamente um obstáculo na notação de PO.

Um modelo causal é composto dos seguintes elementos (simplificados): (1) um critério para a comprovação de validade de uma relação causal, (2) um critério para a comprovação de direcionalidade de uma relação causal, e (3) uma linguagem matemática para expressar relações de causalidade. Os primeiros dois elementos podem ser conectados dentro da ideia de abordagens causais. A linguagem é o que permite transformar abordagens em modelos. Discutiremos esses elementos por vez.

O problema fundamental da causalidade é a conexão entre uma suposta causa e um efeito observado. O que veio a se chamar de Método Científico nada mais é do que o conjunto de práticas e abordagens que conferem um grau aceitável de validade a uma conexão postulada em formato de hipótese. Na sua revisão de tipos de causalidade, Brady (2008) discute quatro principais abordagens<sup>4</sup>: (1) associativa, (2) mecanística, (3) de manipulação e (4) contrafactual. Essa tipologia abarca as principais formas pelas quais a Ciência abordou (e aborda) o problema. Cada uma ilumina um aspecto importante para a comprovação de uma relação causal.

Segundo a abordagem associativa, dois elementos estão causalmente conectados quando é possível repetidamente observar a associação entre eles. Um exemplo simples: toda vez que o sol está prestes a nascer, o galo canta. Um exemplo complexo: quando um ditador personalista morre, a sucessão tende a ser muito violenta (Geddes *et al*, 2018). O argumento subjacente é: se dois elementos estão conectados causalmente, sempre que a causa for observada, é preciso observar o efeito também. A abordagem mecanística enfatiza as **sub-conexões** entre dois elementos. Segundo essa perspectiva, a conexão causal depende da existência de mecanismos entre causa e efeito. No exemplo da morte do ditador personalista, seria possível observar que a ausência de um poder institucionalizado (como um parlamento ou conselho) permite que a morte do ditador produza violência na sucessão. Nesse caso, basta aplicar a lógica associativa também ao mecanismo. Se as associações se sustentam, a relação é causal.

Por sua vez, a abordagem de manipulação afirma a necessidade de **alteração** do ambiente para que se possa observar efeitos. Essa abordagem foi (e continua sendo) muito prevalente nas Ciências Médicas, por exemplo. Ignaz Semmelweis a utilizou extensamente para

---

<sup>4</sup> Essa tipologia não é exaustiva, mas é possível argumentar que modelos modernos de causalidade derivam de uma ou uma combinação dessas abordagens iniciais. Para exemplificar, o modelo combinatorial (Ragin e Pennings, 2005), apesar de possuir uma linguagem causal própria (baseada na teoria dos conjuntos e matemática Booleana), não é discutido neste artigo, mas tem claras origens em uma abordagem associativa com elementos contrafactuais.

descobrir que a ausência da lavagem de mãos produzia febre puerperal em grávidas. Ele primeiramente vetou a ida do padre ao hospital, desconfiando que causava terror nas demais grávidas. Percebendo que essa manipulação (ou **intervenção**) não modificou as taxas de mortalidade, Semmelweis eventualmente requisitou que todos os médicos lavassem as mãos antes de realizarem partos. Essa manipulação o permitiu observar que a taxa de mortalidade caiu, o que era evidência da relação causal entre higiene médica e a ocorrência da febre.

A abordagem contrafactual é, a princípio, a menos concreta das quatro. Nela, uma causa é observada enquanto a diferença entre resultados de dois mundos quase idênticos, sendo um deles factual e outro contrafactual. Retomando o exemplo do ditador: se fosse possível observar o mesmo país com o ditador morrendo e não morrendo ao mesmo tempo, a diferença de níveis de violência entre eles poderia ser atribuída à morte. Em um nível abstrato, isso faz sentido: se a unidade de análise é exatamente a mesma nos dois casos, e tudo de diferente entre eles é uma única variável (ou níveis de uma variável), qualquer diferença de resultados só pode ser atribuída a essa diferença de causas. De um ponto de vista conceitual, essa abordagem é livre de empecilhos, diferentemente das demais. Sempre que se observa diferenças de resultados entre um mundo factual e contrafactual que diferem em uma única causa, não há erros de atribuição da relação causal, ao contrário do que pode ocorrer ao se adotar qualquer outra abordagem. O problema é que - por definição - é impossível observar o mundo contrafactual, algo conhecido como o problema fundamental da inferência causal (Holland, 1986). A consequência de se adotar uma abordagem contrafactual é que a relação de causalidade é tanto mais válida quanto mais o pesquisador se aproxima da comparação contrafactual impossível.

Os dois modelos de causalidade que discutimos aqui importam elementos dessas abordagens, mas aplicam uma linguagem causal própria, além de um tratamento mais rigoroso ao problema da identificação. Adicionalmente, ambos reforçam a conclusão de que toda inferência causal depende dos pressupostos invocados no desenho de pesquisa específico, como detalharemos adiante.

### ***Potential Outcomes***

Nesta seção abordaremos as características principais do modelo causal que possui maior aderência nas ciências sociais: Potential Outcomes (PO). Este modelo também pode ser chamado de Rubin's Causal Model (HOLLAND 1986), ou ainda, de modelo NRH representando

as iniciais de Neyman, Rubin e Holland. Estes são os autores que contribuíram ativamente para a concepção atual do PO (NEYMAN 1935 ; RUBIN 1974, 1977, 1978 ; HOLLAND 1986).

De maneira geral, o modelo de PO trata a identificação do efeito causal a partir da diferença dos resultados potenciais das unidades expostas a tratamentos diferentes. Isto é, diferentes níveis de uma determinada causa estão associados a diferentes níveis de efeitos. A ideia é que a exposição a um certo nível desta causa culminará em um resultado potencial, ao variar o nível de exposição (ou até mesmo na ausência dela) será observado outros resultados potenciais. A identificação do efeito causal está diretamente ligada a comparação desses resultados.

Para facilitar o entendimento do modelo, utilizaremos a lógica de experimentos aleatórios para explicar os conceitos básicos do PO. Ou seja, apesar do modelo abranger causas com natureza contínua, para fins de simplicidade, apenas a forma binária (ausência ou presença) será explorada. A forma mais simples do modelo envolve apenas dois componentes. O primeiro é a variável que indica se a unidade (u) será exposta ao tratamento (t) ou ao controle (c) que chamaremos aqui de S. O segundo componente é o resultado potencial (Y) da unidade que pode ser exposta a condição de tratamento ou de controle. Então, a unidade exposta ao tratamento tem um resultado potencial  $Y_u(1)$ , caso fosse exposta a c, teria o resultado potencial  $Y_u(0)$ . O efeito causal do tratamento em relação ao tratamento controle é dado pela equação a seguir:

$$T = Y_u(1) - Y_u(0)$$

De forma mais simples, ao compararmos uma unidade na condição de tratamento e na condição de controle, a diferença de seus resultados potenciais seria o efeito causal do tratamento. Essa é a equação que dá sentido ao modelo de Potential Outcomes. Ela, no entanto, representa o que Holland (1986) chamou de problema fundamental da inferência. Para identificarmos o efeito do tratamento, precisamos observar dois resultados potenciais que são impossíveis de ser obtidos em simultâneo. Isto é, observar a mesma unidade (u), na condição de tratamento e de controle ao mesmo tempo é impossível. Pensando no exemplo mencionado na Introdução, o médico Semmelweis precisava testar a hipótese de que os hábitos de higiene dos médicos causavam a febre puerperal nas mulheres recém paridas. De acordo com o modelo causal exposto, Semmelweis precisava observar um parto em que o médico higienizou perfeitamente suas mãos (unidade tratada) e observar se a mulher contraiu a febre (resultado potencial). Isso seria o primeiro componente da equação acima. O segundo componente seria o

Os bebês  
foram  
paridos



contrafactual do primeiro: Semmelweis precisaria observar o mesmo médico, a mesma paciente, na mesma cirurgia, na condição de controle, isto é, sem a higienização das mãos (unidade não tratada). Fica claro que nem Semmelweis, nem ninguém, pode observar uma mesma unidade na condição de controle e tratamento ao mesmo tempo.

Nessa perspectiva, seria impossível falar sobre causalção e todo conhecimento produzido até hoje não passariam de meras relações. Não é bem assim. Em seu artigo, Holland (1986), definiu duas soluções para o problema fundamental da inferência: científica e estatística. Como o próprio nome sugere, a primeira está associada a ambientes muito controlados e está sujeita a pressupostos não testáveis. No exemplo citado na introdução sobre a relação entre força aplicada a um objeto e a sua distância percorrida, estamos tratando de uma comparação semelhante a um experimento. Isto é, o estado de controle é a aplicação sobre o objeto de uma força (F) igual a zero, já o estado de tratamento é  $F = X$ , em que X representa a quantidade de força. Os resultados potenciais são o quanto o objeto se locomoveu no estado de controle ( $F = 0$ ) e de tratamento ( $F = X$ ). O efeito de F seria igual a  $(Y|F = X) - (Y|F = 0)$ . Aqui, Y representa a distância percorrida. O problema fundamental da inferência seria que não podemos observar a distância percorrida do objeto nas duas condições ao mesmo tempo. O cientista, no entanto, pode se valer do pressuposto de que o objeto tratado no estado t-1 (antes da aplicação do tratamento) é igual ao objeto no estado de controle ( $Y|F = 0$ ). Com esse pressuposto, o problema está resolvido, pois ambas partes da equação são observáveis em espaços de tempo diferentes. Supondo a distância percorrida do objeto no estado de controle seja 0 e no tratamento seja d, o efeito causal da força aplicada X é:  $T = (Y_t|F = X) - (Y_{t-1}|F = 0) = d - 0 = d$ .

A solução estatística é a que mais interessa às Ciências Sociais. Isso porque a primeira (científica) precisa de pressupostos que raramente podemos postular. O fenômeno social tende a ser mais complexo do que fenômenos naturais, o que torna difícil defender a homogeneidade das unidades no período antes e depois do tratamento (como no exemplo acima). A solução estatística expande o nível de identificação do efeito causal do indivíduo para a média da população. Ou seja, dado uma população U, o pesquisador pode extrair duas amostras similares, aplicar o tratamento para uma delas e comparar o resultado potencial médio do grupo tratado e do grupo de controle. Com isso obtemos a seguinte equação:

$$ATE = E(Y|S = t) - E(Y|S = c)$$

$$ATE = E(Y|S = t) - E(Y|S = c)$$

dada

Nessa equação, trocamos o efeito do tratamento (T) pelo efeito médio do tratamento (ATE), bem como substituímos o resultado potencial do indivíduo Y pelo resultado potencial médio  $E(Y|S)$ . Para que essa equação se sustente, assim como a solução científica, precisamos provar que os dois grupos extraídos da população são comparáveis. Isto é, os grupos de tratamento e controle precisam ser idênticos, exceto pelo fato de um deles ser tratado. Deste modo, a subtração dos resultados potenciais resultará no efeito causal médio do tratamento. O desenho que melhor se acomoda a esse modelo causal são os experimentos aleatórios controlados (Randomized Control Trials). O pressuposto da semelhança entre os grupos é assegurado pelo random assignment, isto é, a distribuição aleatória do tratamento. Por ser aleatório, a condição de tratado não está relacionada a nenhuma característica (observável ou não) dos indivíduos. Esse pressuposto pode ser testado no nível das variáveis observáveis ao comparar os dois grupos no período anterior ao tratamento.

Essa lógica de causalidade também pode ser aplicada em pesquisas cuja fonte de dados seja observacional. Os desenhos quase-experimentais, conhecidos por emular a lógica experimental, aderem perfeitamente ao modelo de PO. A regressão descontínua, por exemplo, contrasta grupos de tratamento e controle comparáveis ao observar unidades que se aproximam de um limite que representa uma descontinuidade na distribuição do tratamento (que divide as unidades em tratadas e não tratadas). Ao comparar indivíduos que estão próximos do ponto de corte, se invoca o pressuposto de que a condição de tratado foi distribuída como se fosse aleatória. Nessa conjuntura pode-se observar o *efeito local médio do tratamento* pela seguinte equação:  $LATE = \lim_{x \rightarrow c} E(Y|X = x) - \lim_{x < c} E(Y|X = x)$ . Em que X corresponde à variável que determina a condição de tratamento. Além da regressão descontínua, outros quase-experimentos, como variáveis instrumentais, pareamento, diferenças-em-diferenças e controle sintéticos são largamente utilizadas na Ciências Sociais e utilizam a lógica de PO para identificação causal.

Para calcular o efeito causal no modelo de *Potential Outcomes*, alguns pressupostos são necessários. Os dois principais são: *Unit Homogeneity* e *Conditional Mean Independence*. Trataremos deles a seguir.

### *Unit Homogeneity*

A homogeneidade das unidades é particularmente necessária para mensurar o efeito causal do indivíduo. Dada a impossibilidade de observar um contrafactual perfeito para unidade  $u$ , o pesquisador precisaria achar outra unidade semelhante à primeira para que esta seja o contrafactual dela. No caso da solução científica ao problema fundamental da inferência, esse pressuposto precisa ser invocado, apesar de não poder ser testado ao nível das variáveis não observáveis. Utilizando mais uma vez o exemplo do médico Semmelweis, para testar sua hipótese observando dois indivíduos comparáveis. Isto é, dado que ele observa um médico que não tenha cuidados de higiene, basta encontrar uma unidade que seja comparável a primeira exceto pelo fato variável de interesse (higiene) e observar o resultado potencial de ambos casos (febre puerperal). Normalmente, esse pressuposto é invocado pelos cientistas naturais, pois em um laboratório controlado e com unidades menos complexas do que seres humanos, a homogeneidade das unidades é mais plausível. Então, tudo que o cientista precisa provar é que  $Y_{u_1}(1) = Y_{u_2}(1)$  e que  $Y_{u_1}(0) = Y_{u_2}(0)$ . Podemos então dizer que  $ITE = Y_{u_1}(1) - Y_{u_2}(0)$ . Agora, ambas partes da equação são observáveis.

"As formas **AMBOS OS** e **AMBAS AS** são usadas quando estiverem seguidas de substantivo."

#### *Conditional Mean Independence*

O segundo pressuposto é equivalente à homogeneidade das unidades para a solução estatística do problema fundamental da inferência. Ou seja, precisa-se supor que, na média, os grupos de controle e de tratamento são independentes do tratamento. Como dito acima, não se pode identificar causalidade a partir de uma (ou duas) unidades apenas quando o pressuposto da homogeneidade não é sustentável. Precisa-se então definir uma população bem definida  $U$ , e a partir dela, extrair duas amostras para compor o grupo de tratamento e o grupo de controle. A escolha de quem vai compor cada grupo precisa ser independente de qualquer pré-característica dos indivíduos. Como dito acima, em um experimento aleatório, esse pressuposto é validado a partir da distribuição aleatória do tratamento (*random assignment*).

Em termos matemáticos, precisa-se provar que  $E(Y_u|S = t) = E(Y|S = t)$  e, por consequência,  $E(Y_u|S = c) = E(Y|S = c)$ . Sendo  $E(Y_u|S = t)$  o resultado potencial de todas as unidades  $u$  que compõem a população  $U$  e a  $E(Y|S = t)$  o resultado potencial da amostra designada para o grupo de tratamento. A mesma lógica é aplicada para o grupo de controle. Não podemos observar  $E(Y_u|S = t)$  e  $E(Y_u|S = c)$  ao mesmo tempo, contudo, aplicando o pressuposto do *conditional mean independence* podemos acessar o efeito causal médio a partir

Frase incompleta

ui (com i subscrito)?

da equação exposta acima em que todas as partes são observáveis:  $ATE = E(Y|S = t) - E(Y|S = c)$ .

### Vieses

Esses pressupostos podem ser violados de diversas formas que culminam em tipos de vieses distintos. Em pesquisas experimentais, a atribuição aleatória do tratamento consegue resolver a maior parte dos problemas na identificação causal, porém nem sempre o pesquisador dispõe das ferramentas necessárias, ou o desenho é impossível de se aplicar. As principais formas de viés são: *baseline bias*, *self-selection bias*, *post-treatment bias* e *non-independence bias* (Przeworski, 2009).

O *baseline bias* ocorre quando os grupos de controle e tratamento diferem para além do tratamento. São diferentes antes mesmo da sua manipulação. Normalmente esse viés ocorre quando a atribuição do tratamento não é aleatória e está correlacionada com alguma variável omitida ( $W$ ). Ou seja, a condição de tratado estará presente apenas em indivíduos em que  $W$  está presente. Logo, o efeito causal estimado seria  $E(Y|S = t, W = 1) - E(Y|S = c, W = 0)$ . Esse valor é enviesado pois o pesquisador não conseguiria distinguir o que é efeito do tratamento e o que é efeito da variável omitida.

O viés de auto-seleção é uma derivação do *baseline bias*. Ele ocorre quando as unidades da população  $U$  identificam alguma vantagem em serem tratadas e buscam o tratamento por si. Voltando ao exemplo de Semmelweis, digamos que ao invés de manipular o tratamento (higiene das mãos) para grupos de tratamento e controle, ele simplesmente observasse os médicos que higienizavam suas mãos e comparasse o resultado com os que não o fazem. Ou seja, a unidade escolhe ser tratada ou não. Se, por alguma razão, existir alguma relação entre higienizar as mãos e uma variável omitida (e.g. qualidade da formação do médico) e que essa variável omitida também esteja correlacionada a obtenção de febre puerperal, a comparação entre os grupos estaria enviesada.

O *post-treatment bias* está relacionado a mudança de atributos da unidade quando o tratamento é aplicado. Deste modo, não podemos distinguir o que é efeito do tratamento e o que é efeito das variáveis que mudaram. Por essa razão alguns atributos do indivíduo não podem ser causa pois estão intimamente associados a diversas características que mudarão quando o seu status muda de controle para tratamento. A raça é um exemplo de atributo que não pode ser causa em um modelo de PO (Holland, 1986). Suponha que podemos manipular a

Não seria o caso de traduzir os termos e colocar os originais em inglês em nota de rodapé?

"obter" costuma ter conotação positiva

seria melhor exemplo de questionários aplicados por período longo o suficiente para ocorrência de mudanças importantes no contexto social.

exemplo  
muito  
absurdo

raça e observar a mesma unidade sendo branca e negra ao mesmo tempo. Após a aplicação do tratamento (raça) a unidade não seria mais ela mesma. Com a raça, diversas variáveis (observáveis e não observáveis) também mudariam.

Por último, o *non-independence bias* acontece quando existe um processo de aprendizagem do grupo controle ao observar o grupo de tratamento. O resultado potencial das unidades não tratadas estaria sendo superestimado (ou subestimado) e o efeito causal estaria enviesado. Imagine que um pesquisador esteja tentando observar o efeito de atividades adicionais no desempenho dos alunos. Para tanto, ele precisaria observar um grupo de estudantes que recebem as atividades e outro que não recebem. Mesmo que a atribuição do tratamento tenha sido aleatória e os dois grupos sejam comparáveis no momento anterior a aplicação do tratamento, se os alunos tratados tiverem contato com os não tratados e repassarem a atividade (ou seu conteúdo), o resultado potencial do grupo de controle seria superestimado e o efeito enviesado.

### **Structural Causal Models e DAGs**

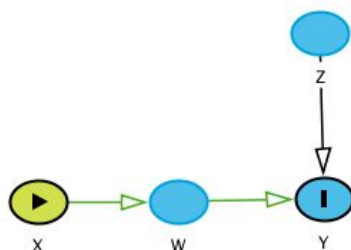
Structural Causal Models se originaram fora das Ciências Sociais e estão relacionados aos trabalhos de Sewall Wright (1921, 1934, 1960) nas áreas de Genética e Estatística. Esses modelos iniciais migraram para as Ciências Sociais sob o nome de Structural Equation Models (Bollen *et al*, 2010), os quais postulavam uma série de funções entre pares ou conjuntos de variáveis. Eventualmente, tanto a técnica de *path analysis* de Wright quanto modelos de equações estruturais caíram em desuso.

Na área da Ciência da Informação, *path analysis* foi modificada para se transformar em um framework completo, com linguagem própria, de causalidade. Essa abordagem recebeu o nome de Structural Causal Models. Sua origem é melhor atribuída aos trabalhos de Judea Pearl (2009). Assim como a lógica de Potential Outcomes, esse modelo parte de uma concepção contrafactual de causalidade acrescida de elementos de manipulação, associação e mecanismos. Causalidade é expressa em **uma intervenção contrafactual nos valores assumidos por uma variável independente**. A presença de efeitos da intervenção denota a relação causal. Ao contrário do modelo NRH, essa relação é expressa através de um operador próprio (*do*), de modo que o efeito causal de X sobre Y pode ser escrito da seguinte forma:

$$E(Y | do(X = 1)) - E(Y | do(X = 0))$$

Como a situação representada por (*do*) é inobservável, o desafio é expressar essa mesma relação em termos de probabilidades condicionais. Para isso, antes de se invocar um ou mais pressupostos (e.g. exogeneidade condicional), se invoca um diagrama causal. Esse diagrama é não-paramétrico e pode ser expresso graficamente ou em termos notacionais. Seu objetivo é documentar as variáveis relevantes para a explicação do fenômeno de interesse e as relações propostas entre elas. Graficamente, essas relações são codificadas na presença ou ausência de setas entre as variáveis e na estrutura de ancestralidade e descendência, como é ilustrado abaixo:

Figura 1 - DAG



Nessa figura, X causa W (devido à seta saindo de X que aponta para W), assim como W causa Y e Z causa Y. Assim W é filho de X, Y é filho de W, Z é pai de Y, e X é um ancestral de Y. Cada par de variáveis conectado por uma seta unidirecional está em uma relação causal incondicional<sup>5</sup>. Essa estrutura causal nos informa que não há empecilhos à estimação do efeito de X sobre Y (entraremos em detalhes adiante). Além disso, poderíamos dizer que W é um **mediador** desse efeito. Um mediador é sempre identificável por estar no caminho causal entre uma variável independente e uma dependente.

Para fins didáticos, é útil pensar em um diagrama causal como um sistema de encanamento pelo qual passa água. Na figura 1, a água flui tanto de X quanto de Z. A partir de X, a água chega a W e depois a Y. A partir de Z, a água chega apenas a Y. Um caminho causal está incondicionalmente aberto desde que haja uma continuidade de setas **na mesma direção** entre uma sequência de variáveis. No jargão da área, X e Y estão *d-connected*, mas o caminho de X a Z está fechado em Y, assim como o caminho de Z a qualquer variável que não seja Y.

<sup>5</sup> Assumindo que o DAG remete a relações de causalidade. É possível desenhar DAGs que representem outros tipos de associações.

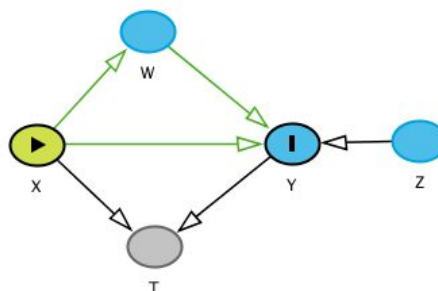
Isso significa que X e Z são ortogonais, ou estão incondicionalmente separados, ou *d-separated*.

Como pode ser observado a partir da figura 2, as relações entre variáveis no diagrama estão todas expressas em termos de relações condicionais, ou seja, para observar o efeito de X sobre Y é preciso assumir que  $P(Y|do(X))$  é igual a  $P(Y|X)$ . Isso só é possível através da observação das probabilidades condicionais informadas pelo diagrama e a verificação do critério de *back-door*, que discutiremos adiante. Porém, para completar a explicação: variáveis cuja conexão com Y está bloqueada por X podem ser desconsideradas na estimação do efeito causal, desde que essas variáveis não possuam setas que direta ou indiretamente apontam tanto para X quanto para Y. apontem

A primeira vantagem de pensar o desenho de pesquisa em termos de um diagrama causal é a obrigatoriedade do próprio diagrama. Essa condição exige que o pesquisador codifique expectativas teóricas, pressupostos e intuição em uma única estrutura cujos elementos são facilmente apreendidos por outros pesquisadores. Cada nó do diagrama, conexão (ou ausência de conexão) entre eles e estrutura proposta são pressupostos do modelo. Assim, avaliar a plausibilidade do desenho de pesquisa se torna imediatamente mais fácil e direta.

A segunda vantagem versa sobre a identificação de controles válidos para análise. Existe hoje uma série de artigos e manuais aplicados à questão de pelo que controlar (Becker, 2005; Greenland e Pearl, 2011; Keele *et al*, 2019; Hünermund e Bareinboim, 2020). Devido às limitações de espaço, nos atemos a duas situações de interesse, ambas codificadas na figura 2.

**Figura 2 - DAG - Controles**



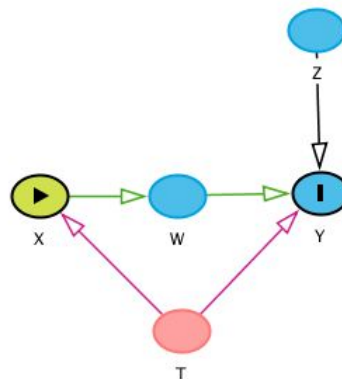
Um pesquisador, sabendo vagamente que as variáveis abaixo são importantes para a explicação de Y pode se sentir tentado a dois erros: controlar por W ou controlar por T. No

diagrama, X e Y estão *d-connected* (incondicionalmente conectados) visto que há um caminho aberto entre eles. Porém (e retomando a questão didática do sistema de encanamento), controlar por W é equivalente a fechar parte da passagem entre eles. Em outras palavras, X e Y são parcialmente **separados** condicionalmente a W. Como W representa um percentual desconhecido da variação de X sobre Y, controlar por W significa filtrar a variação, o que leva à produção de erros de tipo M<sup>6</sup>.

Controlar por T também é um problema. No diagrama, o caminho entre X e Y está fechado em T. Chamamos a variável T de *collider* ( $X \rightarrow T \leftarrow Y$ ). Um *collider* sempre fecha um caminho causal, de modo que se for a única conexão entre duas variáveis, elas devem ser pensadas como *d-separated* ou ortogonais. Contudo, controlar por T **abre** o caminho causal entre X e Y, ou seja, X e Y estão conectados condicionalmente a T. O problema é que essa conexão é espúria, visto que a estrutura causal já informa que a única conexão causal entre X e Y é direta ou via W. Controlar por um *collider* leva à produção de erros do tipo S<sup>7</sup>. Se a figura acima fosse traduzida em um modelo de regressão linear em que tanto W quanto T figurassem enquanto controles, o coeficiente do efeito de X sobre Y poderia estar sub ou superestimado e com a chance de ter o sinal trocado.

A terceira vantagem de DAGs diz respeito à visualização de ameaças à identificação causal. Como explicado anteriormente, um modelo de PO elenca *confounders* e causalidade reversa como principais problemas à identificação. Em um DAG, *confounders* são equacionados a um *backdoor path* entre X e Y.

**Figura 3 - DAG - Confounding**



<sup>6</sup> Quando o pesquisador conclui que a magnitude de determinado efeito é grande (pequena) quando na verdade é pequena (grande) (GELMAN e CARLIN, 2014).

<sup>7</sup> Quando o pesquisador conclui que o efeito é positivo (negativo) quando na verdade é negativo (positivo) (GELMAN e CARLIN, 2014).



A figura 3 adiciona a variável T ao modelo. Fica claro que T determina tanto X quanto Y. A regra do *backdoor path* estabelece que enquanto existirem setas apontando para X que também apontam para Y<sup>8</sup> (ou tenham origem em Y), o efeito de X sobre Y estará confundido<sup>9</sup>. Em linguagem notacional, a distribuição  $P(Y|X)$  é determinada por  $P(Y|T)$  e  $P(X|T)$ , de modo que a variação observada em Y pode ser parcial ou inteiramente uma consequência de T, independentemente dos valores assumidos por X. Esse é o caso clássico de *confounding* e é facilmente comunicado em um DAG.

Outro conceito conhecido em Ciências Sociais é causalidade reversa, cuja interpretação pode significar duas coisas: (1) uma relação causal de simultaneidade (i.e. dois elementos causam um ao outro ao mesmo tempo, como os economistas tendem a conceber oferta e demanda), e (2) um efeito de *feedback*, em que X afeta Y, que por consequência afeta uma terceira variável que é causa de X.

SCM permitem resolver e melhor compreender problemas de causalidade reversa. A Figura 4 ilustra essas duas concepções comuns para causalidade reversa. O problema é que esses diagramas são **cíclicos**, de modo que os critérios de eliminação de *confounding*<sup>10</sup> indicariam que as relações causais de interesse não são identificáveis. Porém, como Pearl e Mackenzie (2018) discutem, causalidade reversa é o produto de uma falha de elaboração causal; o pesquisador falha em levar em consideração a estrutura temporal das variáveis. Se ao invés de X e Y, tivéssemos X1, X2 e Y, poderíamos desenhar  $X1 \rightarrow Y \rightarrow X2$ , o que seria modelável com dados em painel onde houvesse autoregressão.

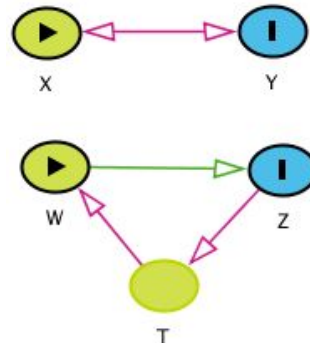
---

<sup>8</sup> Essas setas não precisam apontar diretamente nem para X, nem para Y. Desde que determinada variável tenha conexões causais que eventualmente apontam para X e Y (ou com origem em Y), ela deve ser pensada como um *confounder*.

<sup>9</sup> Ou seja,  $P(Y|do(X))$  não pode ser equacionado a  $P(Y|X)$ , o que indica que é impossível estimar efeitos a partir de dados observáveis.

<sup>10</sup> O que chamamos de *do-calculus*.

**Figura 4 - DAG - Causalidade Reversa**



No caso do segundo diagrama da figura, a mesma lógica poderia ser aplicada. O resultado seria  $W1 \rightarrow \underline{W2} \rightarrow Z \rightarrow T \rightarrow W2$ . A relação causal  $W1 \rightarrow Z$  é identificável, e  $W1$  é apenas um ou mais valores defasados de  $W$ . Relações de simultaneidade também podem ser compreendidas como representações de incerteza em relação a *confounders*. No caso, escrever  $X \leftarrow U \rightarrow Y$  (onde  $U$  é a variável omitida) seria mais proveitoso do que utilizar as setas bidirecionais. O tratamento do problema dado por SCM leva a um questionamento empírico: o quão comuns são relações causais simultâneas e *feedbacks*? Não possuímos uma resposta a essa pergunta.

### **Argumento**

Sendo duas abordagens assentadas em uma lógica contrafactual e preocupadas com *confounding*, seria de se esperar que PO e DAGs dialogassem bastante. Apesar de certa movimentação em prol da integração das abordagens (Hünermund e Bareinboim, 2019), muitos dos principais nomes das Ciências Sociais têm expressado ceticismo a respeito da utilidade de DAGs para o melhoramento da pesquisa social (Imbens, 2019). Em geral, as principais objeções à integração advêm de dois argumentos: os problemas de identificação iluminados por DAGs são igualmente deriváveis da linguagem notacional já existente em PO, e não existem técnicas de identificação (e.g. regressão descontínua) que tenham sido derivadas de abordagens DAG. Nesta seção, argumentamos que, muito pelo contrário, existem diversos pontos de complementaridade entre as duas abordagens.

## **Linguagem Causal e Comunicação de Desenhos**

Felizmente, a avaliação da pesquisa social tem migrado da consideração dos méritos e deméritos de diferentes técnicas de análise (e.g. regressão linear ou logística) e de diferentes tipos de estudo (e.g. qualitativo ou quantitativo) para os elementos dos desenhos de pesquisa.

Seguindo Blair *et al* (2019), pensamos um desenho de pesquisa como uma estrutura que compõe: (1) um modelo de mundo (i.e. uma simplificação teoricamente orientada da realidade), (2) uma hipótese acerca da conexão entre mudanças em uma variável sobre valores assumidos por outra<sup>11</sup>, (3) uma estratégia de coleta de dados e (4) uma técnica de análise desses dados.

Qualquer desenho de pesquisa, por sua vez, é construído com o objetivo de responder uma ou mais perguntas. Em outras palavras, desenhos servem para atingir (ou revelar) determinados **estimandos**, que podem ser pensados como quantidades causais. A capacidade de determinado desenho em atingir um estimando depende da integração de seus elementos e da qualidade de cada um deles (para o estimando específico escolhido pelo pesquisador). Qualidade, é claro, pode possuir um grau desconhecido de subjetividade. Por isso, desenhos tendem a ser avaliados em relação a critérios bem estabelecidos. Um exemplo é o critério de poder, que é a probabilidade do desenho em rejeitar a hipótese nula (i.e. efeito zero). A **precisão** de uma estimativa (distância entre a estimativa e o estimando que representa o “valor real” do efeito) também é comumente utilizada como critério avaliativo. Para pesquisadores de tradição mais qualitativa, um critério bastante comum é *leverage*, que é o quão informativa é a escolha de determinado caso para o estudo.

A relevância do critério depende do objetivo causal do pesquisador. Em uma das pesquisas ilustrativas que discutiremos adiante, o propósito central era saber se o Bolsa Família contribuiu para a reeleição de Lula em 2006. Ou seja, os critérios centrais são a precisão das estimativas, o *Type S rate*, que é a probabilidade de uma estimativa possuir o sinal correto dado que é precisa, e o *Type M rate*, que é a probabilidade da estimativa possuir a magnitude correta dado que o sinal está adequado. Para qualquer pesquisa, existe um número tipicamente grande de critérios avaliativos potencialmente relevantes. Como além da avaliação dos critérios, se avalia também a qualidade de cada elemento individual da pesquisa (e.g. a amostragem do survey foi suficientemente representativa?), avaliar um desenho é uma tarefa difícil, o que

---

<sup>11</sup> É claro que ao invés de uma, pode-se construir uma série de hipóteses com diferentes graus de complexidade (combinações de variáveis, mediação de efeitos, etc). A hipótese pode ser concebida em termos de efeitos de uma intervenção, mas também como condições de necessidade e suficiência.

impõe a necessidade de **comunicação efetiva e eficiente** do desenho. Comunicar dessa maneira significa informar cada elemento do desenho, as perguntas de pesquisa, as quantidades causais estimadas e os testes realizados, tudo isso de maneira concisa e facilmente interpretável.

É aqui que entra o primeiro ponto de complementaridade entre DAGs e PO. Como alguns críticos afirmam, é possível representar qualquer desenho e pressuposto de um DAG em termos de *potential outcomes*. Porém, a **linguagem** de *potential outcomes* tende a ser longa e exigir a representação de uma série de probabilidades condicionais e funções. Já um DAG **linguagem?** permite representar o mesmo desenho graficamente, o que o torna imediatamente mais comunicável, sem a necessidade de traduzir os elementos do desenho em matemática. Todas as probabilidades e pressupostos podem ser facilmente derivados do diagrama causal. Para outros pesquisadores tendo acesso à pesquisa, isso facilita enormemente a árdua tarefa de avaliar um desenho de pesquisa.

### ***Modelo Causal***

Outra vantagem diz respeito à construção do primeiro elemento de um desenho, que é o modelo de mundo ou modelo causal. A abordagem de PO típica começa de uma hipótese e tende a assumir problemas de *confounding* para então criar o desenho que permite razoavelmente eliminar essa endogeneidade potencial. As preocupações com a simplificação da realidade tendem a ser negligenciadas ou discutidas na seção teórica do artigo, sem formalização posterior<sup>12</sup>. Um DAG obriga o pesquisador a pensar no modelo de mundo e a representá-lo com clareza, o que facilita o trabalho de criticar a pesquisa. Potencialmente, essa exigência permite ao pesquisador sair da estreiteza tipicamente relacionada a técnicas derivadas de PO.

Assumir endogeneidade (ou *confounders* desconhecidos) é mais prejudicial do que benéfico à pesquisa científica (Stokes, 2014), do mesmo modo que assumir heterogeneidade de efeitos ou qualquer outro problema *a priori*. Um diagrama causal força o pesquisador a pensar na existência de explicações alternativas (e portanto representar *confounders* no diagrama), na possibilidade de efeitos heterogêneos, na existência de mediadores e todos os demais elementos de relações causais.

---

<sup>12</sup> Algumas são devidamente formalizadas, o que leva a longas representações matemáticas, dando origem ao primeiro tópico discutido nessa seção.

### ***Frontdoor Estimation***

A solução experimental é tipicamente impraticável em grande parte dos estudos em Ciências Sociais, seja por questões éticas, financeiras ou por limitações próprias do experimento<sup>13</sup>. Desenhos observacionais, por sua vez, precisam invocar pressupostos próprios para sustentar inferências, como já discutido. Esses desenhos se tornaram possíveis, em parte, pela lógica de aproximação a uma execução experimental. Uma regressão descontínua, por exemplo, identifica um trecho de determinada distribuição que apresente descontinuidade. O pressuposto é que em uma estreita faixa dos dois lados da descontinuidade unidades tratadas e não tratadas são idênticas (exceto pelo tratamento), porque a seleção para tratamento é como se fosse aleatória. Desenhos de diferenças-em-diferenças assumem que se duas unidades possuem tendências similares ao longo do tempo, mudanças de tendência derivadas de uma intervenção podem ser interpretadas como causais. Variáveis instrumentais identificam uma terceira variável que determina os valores de X sem afetar Y diretamente, e se baseiam no pressuposto da *exclusion restriction*. A multiplicidade de soluções observacionais para o problema de identificação permite ao pesquisador selecionar a que melhor se adequa à estrutura de seus dados.

Cada desenho observacional acompanha limitações próprias, tal qual cada desenho experimental. Em uma regressão descontínua, sempre se identificam efeitos locais (*Local Average Treatment Effects*), que podem ser irrelevantes a depender da hipótese da pesquisa (e.g. uma hipótese que pressupõe um efeito geral ou para extremos de uma distribuição). Diferenças-em-diferenças não lidam com manipulações simultâneas ou tratamentos concomitantes, e variáveis instrumentais podem tanto identificar apenas efeitos localizados (*Average Complier Effects*) quanto identificar efeitos extremos (a depender da natureza do efeito do instrumento sobre a variável independente). Em conclusão, é possível haver situações

---

<sup>13</sup> Apesar da consagração de experimentos, em geral *Randomized Controlled Trials* (RCTs), na pesquisa social (e.g. prêmios Nobel de Economia em 2019), é importante deixar claro que não são balas de prata. Dado que o pesquisador deseja saber um efeito direto entre duas variáveis, não possui problemas éticos ou financeiros, a variável independente é manipulável, validade externa e artificialidade não são preocupações, infrações dos pressupostos inexistem ou estão bem controladas, e o experimento em si não introduza novos vieses, um RCT é o melhor desenho. Considerando o número de condições (e nem mencionamos todas) para uma execução adequada, deveria ser claro que uma parte substancial do mundo social é - pelo menos até agora - intratável experimentalmente.

em que desenhos observacionais existentes sejam inadequados ou ineficientes para identificação causal.

O terceiro ponto de complementaridade entre PO e DAGs diz respeito à contribuição de um desenho de pesquisa observacional adicional, conhecido como *Frontdoor Estimation*. Esse tipo de desenho corrige problemas de endogeneidade ao identificar um mecanismo entre X e Y que seja exógeno a potenciais *confounders*. Isso significa que se pode estimar o efeito de X sobre Y em um processo de dois passos, similarmente a uma regressão com instrumentos. Como o mecanismo M é exógeno aos *confounders*, é possível estimar o efeito de X sobre M e depois de M sobre Y. Esse efeito indireto resultante é robusto à endogeneidade. Bellemare *et al* (2019) demonstram como esse método identifica um efeito causal mesmo na presença de uma multiplicidade de mediadores.

### **Comparabilidade e Generalização**

As complementaridades vão nos dois sentidos. Em um Structural Causal Model (ou DAG), o contrafactual é em termos da intervenção nos valores de X ( $do(X=x)$ ); já no framework de PO, o contrafactual é discutido em termos da comparabilidade entre grupos. Assim, pesquisadores na linha de PO possuem uma atenção especial a o **que** será comparado. É importante lembrar que um efeito causal precisa ser observado em perspectiva comparada, dado que é preciso observar a distribuição de Y em um mundo onde a intervenção em X não ocorreu. Essa lógica não está ausente em DAGs, porém o foco é na construção e derivações dos diagramas. A necessidade de discutir a composição dos grupos sendo comparados é uma força do framework de PO e deu origem a uma série de técnicas de pesquisa específicas (Donald and Lang, 2007; Abadie *et al*, 2014).

A atenção nas composições dos grupos e no potencial de comparabilidade inspirou a investigação de uma série de desvios ou dificuldades de estimação do ATE, a exemplo de problemas de contaminação, *defiers* e *self-selection* (Gerber e Green, 2012). Dados esses (e muitos outros) problemas, o pesquisador pode ainda estimar outros efeitos, a exemplo do *Average Treatment Effect on the Treated* (ATT), *Intent to Treat Effect* (ITT) e *Local Average Treatment Effect* (LATE), que são quantidades causais contidas no ATE. Esses sub-efeitos representam a capacidade do pesquisador de estimar efeitos para uma população inteira. O framework de PO possui um foco substancial na generalização de resultados e sub-populações, algo que não é imediatamente observado em um DAG.

## ***Indeterminação Teórica***

Um DAG codifica informação prévia sobre determinado outcome de interesse em sua estrutura. Assim, a plausibilidade do diagrama pode ser avaliada de acordo com o conhecimento acumulado a respeito do objeto de estudo. Em situações de expectativas teóricas antagônicas, seria necessário o desenho de dois ou mais DAGs para representar os distintos pressupostos acerca do funcionamento do mundo. Por outro lado, existem fenômenos em que é possível um número substancial de expectativas, ou não existe fundamentação teórica bem desenvolvida, ou ainda a estrutura causal evidencia a presença de comportamento cíclico (i.e. causalidade reversa ou *loops*). Se o desenvolvimento teórico é insuficiente ou incapaz de lidar com a natureza cíclica da relação causal, existe pouco que as ferramentas de um DAG possam fazer. Aqui entra a última complementaridade entre PO e DAG. Um modelo de PO sempre assume *confounders* não-observáveis, de modo que o *toolkit* de técnicas de análise já está paramentado para encontrar e explorar quasi-experimentos onde a atribuição do tratamento é aleatória ou próxima disso, o que permite estimar efeitos causais que na estrutura causal geral seriam impossíveis.

## **Metodologia**

A parte empírica deste artigo possui dois propósitos: o primeiro é demonstrar a distância substancial entre a pesquisa social nacional e os frameworks discutidos. Não queremos afirmar que pesquisas acadêmicas estejam sendo feitas sem técnica (apesar de algumas de fato serem), mas que parte da produção de conhecimento científico ainda está sendo conduzida sem atenção às questões elementares que substanciam inferências causais, o que dificulta confiar nos resultados. Nosso segundo propósito é didático. Apresentaremos dois artigos empíricos de qualidade, publicados em revistas internacionais, e demonstraremos como cada pesquisa é traduzida nas linguagens de PO e DAGs. No processo, esperamos evidenciar como esses frameworks auxiliam na sumarização e avaliação dos desenhos propostos, e como eles são capazes de revelar problemas ou potenciais fragilidades.

Para o primeiro propósito, coletamos sistematicamente... [PRESENTE NA VERSÃO FINAL]. **O artigo já está muito longo e poderia ser dividido em três ou quatro:**

- 1. PO**
- 2. DAG**
- 3. Complementaridade entre PO e DAG**
- 4. Avaliação de pesquisa nacional**

Em relação ao segundo propósito, selecionamos dois artigos de Ciência Política, um da área de sucesso eleitoral e outro da área de representação feminina. A escolha dessas pesquisas seguiu três critérios: (1) o número de citações do trabalho, (2) a temática, e (3) o *leverage* de cada escolha (i.e. o quão informativa cada pesquisa é para o propósito de ilustrar os conceitos apresentados no artigo). Como o número de citações possui um forte componente temporal (trabalhos mais antigos são mais citados simplesmente por estarem a mais tempo no mercado acadêmico), selecionamos um dos artigos com o critério de ter sido publicado a partir de 2015.

O primeiro artigo é *The President's New Constituency: Lula and the Pragmatic Vote in Brazil's 2006 Presidential Elections* (Zucco, 2008). Dentro da área de estudos eleitorais na América Latina, esse é o quarto artigo mais citado. O segundo artigo é *Presidentas Rise: Consequences for Women in Cabinets?* (Reyes-Housholder, 2016). Dentre artigos de Ciência Política em revistas da América Latina e entre 2015-2020, esse também é o quarto mais citado. Nós apresentaremos a pergunta de pesquisa para cada artigo, a hipótese central e um resumo da metodologia e resultados. Em seguida, faremos uma representação dos desenhos mobilizados por cada autor, os pressupostos subjacentes e potenciais fragilidades inferenciais a partir dos dois frameworks discutidos.

## **Resultados**

### ***Preocupação Causal nas Ciências Sociais Brasileiras***

[EM CONSTRUÇÃO]

#### ***Artigos Ilustrativos***

Para ilustrar os ganhos de se adotar uma abordagem empírica explicitamente embasada em frameworks causais, faremos uma análise de dois artigos de Ciência Política sobre o Brasil. A seleção desses artigos não deriva de nossas avaliações acerca do comprometimento dos autores com algum dos frameworks apresentados aqui. Buscamos pesquisas em áreas com as quais temos familiaridade, de pesquisadores que, aos nossos olhos, são competentes e rigorosos em suas produções, e nosso objetivo maior é encontrar situações que nos permitam



ilustrar os requerimentos de uma análise causal embasada na complementaridade entre dois frameworks causais.

O primeiro artigo é *The President's New Constituency: Lula and the Pragmatic Vote in Brazil's 2006 Presidential Elections* (Zucco, 2008). Nele, o objetivo central do autor é saber se o Bolsa Família contribuiu para as chances de reeleição de Lula em 2006. A hipótese é que sim. Zucco (2008) argumenta que os municípios mais pobres e vulneráveis no Brasil possuem um comportamento pragmático e distintamente relacionado ao voto econômico que os faz tipicamente apoiar o governo de ocasião<sup>14</sup>, o que foi reforçado pela criação e expansão do Bolsa Família durante o primeiro mandato de Lula.

Temos então uma relação causal direta (Bolsa Família -> Voto), uma relação causal direta para uma subpopulação específica (municípios vulneráveis) e duas outras relações causais: (município vulnerável -> voto no incumbente) e (incumbência -> perfil sociodemográfico dos eleitores do PT). É importante mencionar que o autor se propõe a testar rigorosamente apenas a primeira relação causal. Para isso, Zucco (2008) emprega um estudo observacional em que a unidade de análise são municípios brasileiros, a variável independente é o percentual da população que recebeu o Bolsa Família antes da eleição, e a variável dependente é o percentual de votos dados a Lula agregados por município. O autor emprega regressões lineares com a adição de 10 controles. A tabela completa com o resumo da pesquisa pode ser encontrada no apêndice. O autor encontra resultados favoráveis às hipóteses.

A partir de uma lógica de PO, a questão central é verificar se podemos garantir *ignorability* entre tratamento (recebimento do bolsa família) e voto no candidato Lula em 2006. Uma preocupação relacionada é que o mecanismo de atribuição do Bolsa Família (as condições econômicas da população) também afete a propensão a votar em Lula. O autor emprega um estudo com uma série de controles, alguns dos quais podem ser compreendidos como *proxies* para as condições econômicas do município, a exemplo do PIB. Apesar de não estar explicitado no artigo, a função dos controles em uma regressão é justificar um pressuposto de exogeneidade condicional, o que chamamos de seleções em observáveis. Assume-se que dados os controles a variável independente é exógena (termo de erro não correlacionado com o tratamento) à dependente. Na literatura de PO, essa pressuposição é considerada forte, ou seja, menos plausível. Esse tipo de estudo está sempre sujeito à presença de *confounders*. Uma forma de contornar o problema com os mesmos dados seria

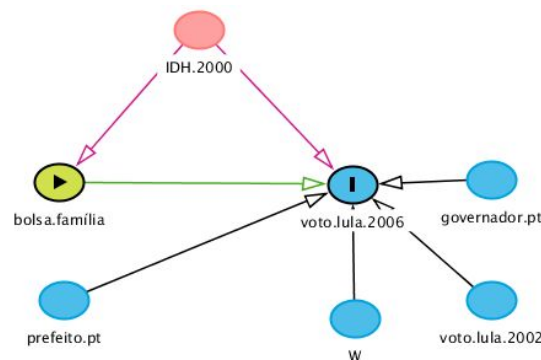
---

<sup>14</sup> O autor produziu uma série de outras pesquisas sobre o assunto (Zucco, 2015; 2020).

conduzir uma análise de sensibilidade, o que poderia indicar a prevalência e magnitude necessárias para um *confounder* hipotético tornar os estimadores indiferenciáveis de zero.

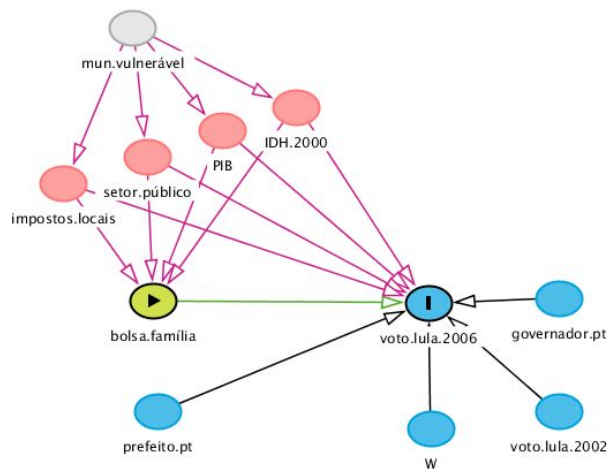
A lógica de SCM começa com um diagrama. Todos os pressupostos precisam estar codificados no diagrama. Reproduzimos aqui nossa avaliação do diagrama causal para a pesquisa de Zucco (2008). Nele, o efeito causal do Bolsa Família é estimável dado que o IDH controla pelo *confounding* produzido pelas características econômicas do município. As variáveis relativas à população do município, percentual de pentecostais e de não-brancos foram colapsadas em W para simplificar o DAG.

**Figura 5 - Zucco - Modelo 1**



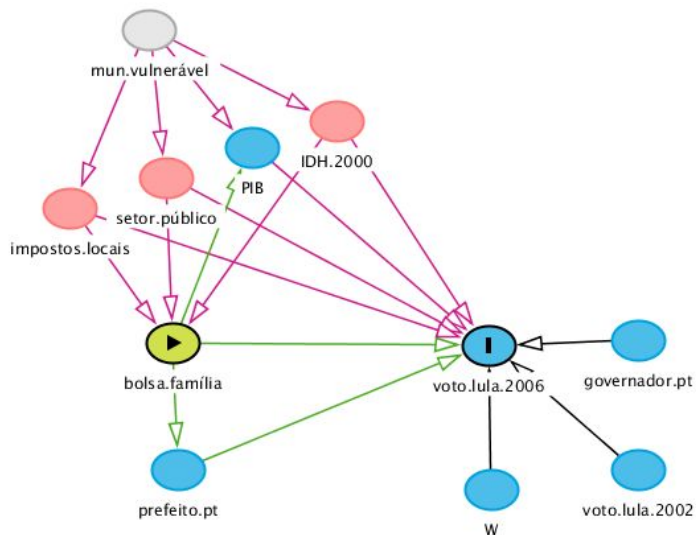
É interessante notar que - exceto o controle do Índice de Desenvolvimento Humano - nenhum dos outros controles é estritamente necessário para a identificação do efeito causal. Nenhum deles integra um *backdoor path* entre X (Bolsa Família) e Y (Voto), de modo que sua inclusão, na melhor das hipóteses, pode ser justificada como melhoria da precisão dos estimadores (Greenland e Pearl, 2011). Em seguida, Zucco (2008) argumenta que municípios vulneráveis são especialmente propensos a apoiar o incumbente, e também são aqueles que possuem maior probabilidade de ter proporcionalmente mais beneficiários do Bolsa Família. Assim, o autor inclui outras variáveis cujo objetivo é representar a subpopulação de municípios vulneráveis. São elas: a capacidade fiscal do município, o tamanho do setor público e o PIB local. A lógica é que essas variáveis capturam o efeito de ser um município vulnerável, o que por sua vez afeta tanto o percentual de recebimento do Bolsa Família quanto o percentual de votos dados a Lula. A figura 6 ilustra esse acréscimo.

**Figura 6 - Zucco - Modelo 2**



Sob esse novo cenário, o efeito causal total ainda é identificável, dado que o *backdoor path* entre *mun.vulnerável*, *bolsa.familia* e *lula.vote.share* está fechado pelo controle dos *proxies* (*PIB*, *impostos.locais*, *setor.público* e *IDH*). Mas esse não é o único diagrama plausível para a pesquisa. Por exemplo, o Bolsa Família foi lançado em 2003, um ano antes das eleições municipais de 2004. Assim, é possível que o programa tenha tido efeito sobre as chances eleitorais de candidatos do PT nas disputas por prefeituras. Adicionalmente, um dos propósitos do bolsa família é melhorar as condições socioeconômicas dos beneficiários, o que pode se traduzir em aumentos marginais do PIB municipal devido ao aumento do consumo dessas famílias.

**Figura 7 - Zucco - Modelo 3**



Nessas condições, o efeito total do Bolsa Família não é estimável devido ao controle por um mediador (ou intermediário) do efeito. Controlar por um mediador filtra parte da variação de Y causada por X. Assim, o único efeito estimável é o efeito direto (não-mediado) do Bolsa Família. Esse não é o alvo causal da pesquisa, mas não invalida os resultados obtidos. Assumindo que apenas o efeito direto foi estimado pela regressão, é plausível supor que o efeito **total** do programa é na verdade **ainda maior** do que o reportado por Zucco (2008). Na verdade, esse parece ser o caso. Comparando os modelos 3 e 4 da tabela 3 do artigo do autor, percebe-se que remover o PIB local aumenta a magnitude do coeficiente de .354 para .363. Nesse DAG, o Bolsa Família possui dois mediadores; para o controle de cada um deles, há um aumento da probabilidade de subestimação do efeito total, também conhecida como erro do tipo M<sup>15</sup>. Potencialmente, a exclusão do PIB e das variáveis relativas ao sucesso dos prefeitos em 2004 seria benéfica à qualidade dos estimadores<sup>16</sup>. Como o autor não reporta um modelo incluindo todas as variáveis com exceção dessas duas, não temos como observar como as magnitudes se comportam.

O exercício feito para o artigo pode ser estendido para variáveis fora do modelo. Por exemplo, se um pesquisador avaliando a pesquisa concebe uma variável que componha um *fork* com as variáveis independente e dependente (bolsa.família <- U -> voto.lula.2006), essa variável pode ser incluída no DAG e testada empiricamente. É também possível conceber variáveis adicionais que sejam *colliders*, por exemplo. Nesses casos, as regras apresentadas neste artigo permitem rapidamente identificar quais são benéficas ou não à identificação causal. Um último ponto: considerando que nossa avaliação a respeito do PIB é verdadeira, em uma análise futura (com dados em painel), o PIB não deveria ser incluído enquanto controle. Isso porque se o Bolsa Família promove ganhos marginais de PIB local, que por sua vez afeta a propensão em votar em Lula em 2006, pode afetar a probabilidade de receber mais transferências no futuro (já que o Bolsa Família é condicionado às condições locais). Assim, o PIB integraria um *backdoor path* entre receber o Bolsa Família até 2006 e receber o Bolsa Família até 2010.

O segundo artigo é *Presidentas Rise: Consequences for Women in Cabinets?* (Reyes-Housholder, 2016). Nele, a autora deseja saber se a eleição de uma presidente causa o apontamento de mais ministras. A hipótese é que sim. Reyes-Housholder (2016) argumenta

---

<sup>15</sup> Gelman e Carlin (2014).

<sup>16</sup> Devido à incerteza sobre o papel do PIB local, o pesquisador poderia decidir manter a variável, assumindo perdas na estimação da magnitude em troca de validade interna.

que presidentes mulheres apontam mais ministras porque mais mulheres integram suas redes sociais. Cada presidente possui um *pool* de candidatos a posições de ministério. A entrada de qualquer pessoa nesse *pool* depende de uma série de características resumidas como capital político, que inclui relações políticas, influência, dentre outras características. Como a presença na rede social do presidente é uma delas, espera-se que exista um percentual maior de mulheres nos *pools* de presidentes mulheres, o que leva a uma maior probabilidade e taxa de apontamento. A autora ainda deriva duas hipóteses secundárias: (1) o apontamento de ministras é maior nos primeiros anos de governo da presidente e (2) estereótipos relativos ao ministério (e.g. sociedades atribuem feminilidade a ministérios como Educação) aumentam a probabilidade da eleição de uma mulher para o cargo.

Temos então uma relação direta (sexo do presidente -> apontamento de ministras) e duas relações **moderadas**: espera-se que o período do mandato e o estereótipo do ministério produzam uma modificação dos efeitos principais, mas que não necessariamente alterem a relação causal. A autora emprega regressões lineares e logísticas onde a proporção de ministras e o sexo da ministra são variáveis dependentes, e o sexo da presidente a independente. Além disso, são utilizados seis controles, incluindo a proporção de ministras durante o mandato do presidente anterior e o percentual de mulheres no Congresso Nacional, ambas variáveis que podem afetar o *pool* de candidatas disponíveis à presidente. A unidade de análise são os ministros. O quadro com o resumo do desenho pode ser encontrado no apêndice

17.

Da mesma forma que a pesquisa anterior, o pressuposto central aqui é exogeneidade condicional. A preocupação também é a mesma: *confounders* não-observáveis. A pesquisa depende de um pressuposto forte, e é ainda mais difícil controlar pelas variáveis que causam a atribuição do tratamento (i.e. o sexo do presidente), devido ao fato de que esse é um exemplo de variável não-manipulável. Segundo alguns autores associados a PO, variáveis não-manipuláveis<sup>18</sup> não podem originar relações causais (CITAR), o que reduz mais a plausibilidade da inferência. A lógica da comparabilidade também se aplica aqui: mulheres estão sujeitas a uma miríade de diferenças em relação aos homens, potencialmente desde a primeira infância. Assim, a não ser que todas essas diferenças sejam capturadas pelas redes

---

<sup>17</sup> Esse é o resumo da parte quantitativa da pesquisa. Existe também um estudo de caso qualitativo da Argentina. Porém, a própria autora informa que a pesquisa qualitativa tem o propósito de ilustrar o argumento ao invés de produzir inferências válidas.

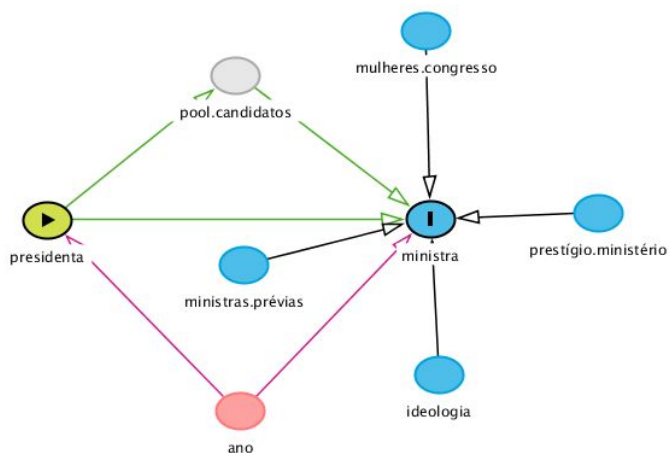
<sup>18</sup> É importante distinguir não-manipulável de não-manipulada. A segunda justifica o uso de um desenho observacional; já a primeira indica uma variável cujos mecanismos de atribuição são incontroláveis.

sociais que as presidentes formam, a comparação entre presidentes homens e presidentes mulheres não seria válida, o que enviesaria todos os estimadores.

A partir de um framework de PO, uma recomendação possível é encontrar subpopulações para as quais o sexo do político eleito seja plausivelmente aleatório (e.g. algum modelo de regressão descontínua<sup>19</sup>). *Mock elections* em experimentos de laboratório poderiam servir de validação adicional: mesmo que esses desenhos ainda não manipulem o sexo, eles controlam (via aleatorização) pelos *confounders* não-observáveis, e podem manipular - dentro de uma amostra de pessoas do mesmo sexo - a eleição a uma posição de poder.

Também criamos o que, em nossa avaliação, é o diagrama causal da pesquisa. Nele, o efeito causal total é estimável dado a ausência de *backdoor paths* entre X (sexo do presidente) e Y (sexo do ministro). A plausibilidade desse diagrama nos parece limitada devido à presença de duas variáveis não consideradas na pesquisa: o sistema partidário e o secularismo<sup>20</sup> da sociedade. Reyes-Housholder (2016) discute a segunda variável, mencionando pesquisas que indicam não ser um preditor forte do apontamento de ministras. Contudo, isso não impede que possua efeitos **indiretos** sobre o sexo. Nesse diagrama, suprimimos as moderações propostas a título de simplificação, nos atendo apenas ao efeito principal<sup>21</sup>.

**Figura 8 - Reyes-Housholder - Modelo 1**



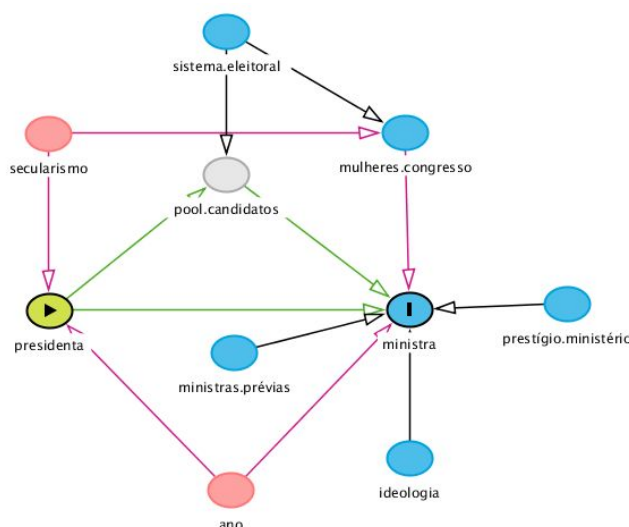
<sup>19</sup> Desde que o argumento seja válido para outros cargos executivos, como prefeitos e governadores.

<sup>20</sup> Estamos utilizando secularismo para definir atitudes pós-materialistas, que tendem a incluir noções de igualdade entre homens e mulheres. Os leitores podem propor as próprias variáveis, como preconceito de gênero estrutural, machismo, etc.

<sup>21</sup> Um moderador (ou modificador de efeitos) pode ser representado como uma variável que afeta um mecanismo (conhecido ou não) entre X e Y. No apêndice é possível encontrar o modelo de Reyes-Housholder (2016) com a inclusão de moderadores.

Dadas as variáveis mobilizadas pela autora, o efeito é identificável. Apenas a variável *ano* é um *confounder*, já que o ano da eleição tanto da presidente quanto do apontamento da ministra podem significar maiores ou menores transformações culturais conducentes à chegada de mulheres ao poder. A Figura 9 ilustra o que queremos dizer com *confounders* indiretos. Também consideramos uma variável ausente nas análises da autora: o sistema partidário. Sistemas partidários mais proporcionais tendem a facilitar a eleição de grupos minoritários ou sub-representados de maneira mais geral (Icite). Assim, é plausível supor que países com sistemas mais proporcionais deem mais chances à eleição de mulheres. Isso significa que o sistema partidário pode afetar o número de mulheres no Congresso Nacional e diretamente a proporção de mulheres que integraria a rede social da presidente<sup>22</sup> (*pool.candidatos*).

**Figura 9 - Reyes-Housholder - Modelo 2**



Esse DAG ilustra nosso argumento sobre complexidade causal. A relação *presidenta*  $\leftarrow$  *secularismo*  $\rightarrow$  *mulheres.congresso*  $\rightarrow$  *ministra* indica que secularismo é um *confounder* indireto da probabilidade de apontamento de uma mulher a um ministério. Nesse caso, controlar pelo secularismo do país ou pelo número de mulheres no Congresso Nacional já bastaria. Porém, o sistema eleitoral também afeta o número de mulheres no Congresso Nacional. Notem que o micro-diagrama *sistema.eleitoral*  $\rightarrow$  *mulheres.congresso*  $\leftarrow$  *secularismo* representa um *fork* invertido que tem um *collider* no número de mulheres no Congresso. Como a autora controla essa variável (o que remove o *confounding* de *secularismo*), ela acaba criando uma associação

<sup>22</sup> A lógica é que sistemas eleitorais mais proporcionais permitem maior participação política feminina (devido a menores custos de entrada e disputa no mercado político), o que permitiria a mais mulheres acumularem capital político e serem incluídas nas redes sociais de presidentes de ambos sexos.

espúria entre *sistema.eleitoral* e *secularismo*. Em si, isso não é problemático, já que nenhuma dessas variáveis é de interesse. Contudo, como o sistema partidário também afeta o mediador *pool.candidatos*, o controle pelo *collider* abre um *backdoor path* entre *presidenta* <- *secularismo* -> *sistema.eleitoral* -> *pool.candidatos* -> *ministra*. Desse modo, nossa avaliação - baseada na plausibilidade do DAG - é que o efeito principal estimado no artigo não possui interpretação causal.

Em tese, seria possível desenhar diagramas ainda mais complexos (e.g. o número prévio de ministras deve afetar a composição das redes sociais de presidentes), porém - dadas as variáveis na figura 9 - as relações propostas já permitem perceber o problema de identificação na estrutura. Diversas soluções são possíveis a partir dessa análise. Dois exemplos: (1) controlar pelo secularismo nacional fecharia qualquer possibilidade de *backdoor path*, (2) encontrar um instrumento para a eleição de uma mulher presidente.

## **Conclusão**      Quando podemos começar a usar o texto em sala de aula?

Frameworks, ou modelos causais, são atualmente o melhor instrumento para pesquisadores identificarem e pensarem sobre relações causais. Esses modelos são sofisticadas concepções contrafactuais com elementos intervencionistas, associativos e de mecanismos. Neste artigo, discutimos o atual modelo dominante nas Ciências Sociais, conhecido como Potential Outcomes (Neyman, 1923; Rubin, 1974; Holland, 1986). Demonstramos como uma relação causal é concebida pelo modelo e como a solução via estimação do ATE permite contornar a não-observabilidade de um contrafactual. Em seguida, apresentamos um modelo cuja concepção mais moderna advém da Ciência da Informação e Epidemiologia (Pearl, 2009), conhecido como *Structural Causal Model* ou DAG. Apesar de compartilhar concepções contrafactuais com PO, SCM emprega uma linguagem causal própria centrada no operador *do* e uma representação gráfica, não-paramétrica, de estruturas causais. Argumentamos que esses dois modelos, longe de opostos, possuem uma série de pontos de complementaridade que se considerados na produção ou avaliação de desenhos de pesquisa potencializam a qualidade inferencial. Ilustramos nossos pontos com a análise de dois artigos empíricos, publicados em periódicos da América Latina, dentre os mais citados em suas respectivas áreas.

Esperamos que essa introdução didática aos frameworks e discussões subjacentes permita e encoraje pesquisadores a incorporar preocupações causais rigorosas em suas



produções. A Ciência compartilha critérios de validação que precisam ser incorporados e discutidos. Além disso, é possível atrelar a baixa credibilidade de algumas pesquisas ao desconhecimento (ou até repúdio) a modelos formais de causalidade. Esperamos que este artigo contribua para a mudança necessária na pesquisa social brasileira.

## Referências

- ABADIE, Alberto, A. D., & Hainmueller, J. (2014). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59 , 495-510.
- ANGRIST, J., & Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24 , 3-30.
- ASHLEY, D., & Orenstein, D. (2005). *Sociological theory: Classical statements*. Boston, MA: Allyn and Bacon.
- BECKER, T. (2005). Potential problems in the statistical control of variables in organizational research: A qualitative analysis with recommendations. *Organizational Research Methods*, 8 , 274-289.
- BELLEMARE, M., & Bloem, J. (2019). The paper of how: Estimating treatment effects using the front-door criterion.
- BLAIR, Graeme, A. C., Jasper Cooper, & Humphreys, M. (2019). Declaring and diagnosing research designs. *American Political Science Review*, 113 , 838-859.
- BRADY, H. (2008). Causation and explanation in social science. In H. B. Robert Goodin Janet Box-Steffensmeier & D. Collier (Eds.), *The oxford handbook of political methodology* (p. 217-270). New York, NY: Oxford University Press.
- BOLLEN, K., S. C., Daniel Bauer, & Edwards, M. (2010). An overview on structural equation models and recent extensions. In L. T. Stanislav Kolenikov & D. Steinley (Eds.), *Statistics in the social sciences: Current methodological developments* (p. 37-73). New Jersey, NJ: John Wiley and Sons.
- DONALD, S., & Lang, K. (n.d.). Inference with difference-in-differences and other panel data. *The Review of Economics and Statistics*, 89 , 221-233.
- DURKHEIM, E.. (2001). *As regras do método sociológico*. Petrópolis, RJ: Editora Vozes.
- GEDDES, Barbara, J. W., & Frantz, E. (2018). *How dictatorship work: Power, personalization and collapse*. New York, NY: Cambridge University Press.

- GELMAN, A., & Carlin, J. (2014). Beyond power calculations: Assessing type s (sign) and type m (magnitude) errors. *Perspectives of Psychological Science*, 9 , 641-651.
- GREENLAND, S., & Pearl, J. (2011). Adjustments and their consequences: Collapsibility analysis using graphical models. *International Statistical Review*, 79 , 401-426.
- HÜNERMUND, P., & Bareinboim, E. (2019). Causal inference and data-fusion in econometrics.
- HÜNERMUND, P., & Louw, B. (2020). On the nuisance of control variables in regression analysis.
- HOLLAND, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81 , 945-960.
- HUME, D. (1739). *A treatise of human nature*. London: John Noon.
- IMBENS, G. (2019). Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics.
- KEELE, L., R. S., & Elwert, F. (2019). The causal interpretation of estimated associations in regression models. *Political Science Research and Methods*, 8 , 1-13.
- MICKEY, R., & Greenland, S. (1989). The impact of confounder selection criteria on effect estimation. *American Journal of Epidemiology*, 129 , 125-137.
- NEYMAN, J., D. D., & Speed, T. (1990). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, 5 , 465-472.
- PEARL, J. (2009). *Causality: Models, reasoning, and inference*. New York, NY: Cambridge University Press.
- PEARL, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. New York, NY: Basic Books.
- PRZEWORSKI, A. (2009). Is the Science of Comparative Politics Possible? In C. Boix & S. Stokes (Eds.), *The Oxford Handbook of Comparative Politics*. Oxford: Oxford University Press
- RAGIN, C., & Pennings, P. (2005). Fuzzy sets and social research. , 33 , 423-430.
- REYES-HOUSHOLDER, C. (2016). Presidentas rise: Consequences for women in cabinets? *Latin American Politics and Society*, 58 , 3-25.
- RUBIN, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66 , 688-701.
- SMITH, W. T. (1856). A course of lectures on the theory and practice of obstetrics. *The Lancet*, 503-505.
- STOKES, S. (2014). A defense of observational research. In D. Teele (Ed.), *Field experiments and their critics* (p. 33-57). New Haven: Yale University Press.

- WRIGHT, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20 , 557-585.
- WRIGHT, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, 5 , 161-215.
- WRIGHT, S. (1960). Path coefficients and path regressions: Alternative or complementary concepts? *Biometrics*, 16 , 189-202.
- WYKTICKY, H., & Skopec, M. (1983). Ignaz phillip semmelweis, the prophet of bacteriology. , 4 , 367-370.
- ZUCCO, C. (2008). The president's 'new' constituency: Lula and the pragmatic vote in brazil's 2006 presidential elections. *Journal of Latin American Studies*, 40 , 29-49.
- ZUCCO, C. (2015). The impacts of conditional cash transfers in four presidential elections. *Brazilian Political Science Review*, 9 , 135-149.
- ZUCCO, C., J. L., & Baykal, G. (2020). Do conditionalities increase support for government transfers? *Journal of Development Studies*, 56 , 527-544.