

Resumo

O artigo apresenta uma introdução didática sobre desenhos de pesquisa, inferência e causalidade em Ciência Política e Relações Internacionais. Destina-se, portanto, a familiarizar estudantes do campo com o ferramental desta atividade científica, bem como introduzi-los aos principais debates que perpassam a disciplina. Para tanto, revisamos os elementos centrais do livro *Designing Social Inquiry* e as publicações que responderam a ele, endossando ou contestando as suas concepções sobre as melhores práticas científicas. Com este artigo esperamos equipar os alunos de Ciência Política e Relações Internacionais para o pensamento crítico sobre as escolhas envolvidas na elaboração e execução de um desenho de pesquisa.

Palavras chave: ciência política, metodologia, inferência, causalidade, KKV

Abstract

This article seeks to present, in a pedagogical and introductory way, central issues concerning research design, inference and causality in Political Science and International Relations. It aims, therefore, at familiarizing students of this field with the common tools of this scientific endeavor, as well as introducing them to the main debates that cut across the discipline. In order to do so, we review the guidelines presented in the book *Designing Social Inquiry* and in the publications that answered it, either endorsing or contesting its conceptions on the best scientific practices. The objective of this contribution is to better equip Political Science and International Relations students to think critically about the choices involved in the elaboration and execution of a research design.

Key words: political science, methodology, inference, causality, KKV

“Agradeço ao(s) parecerista(s) pelas sugestões de revisão do artigo”

Desenho de pesquisa, inferência e causalidade em Ciência Política e Relações Internacionais: Uma introdução didática

Rafael Mesquita
Departamento de Ciência Política - UFPE

Introdução

Uma das principais dificuldades dos praticantes do campo da Ciência Política e Relações Internacionais no Brasil tem sido o domínio apropriado de desenhos de pesquisa e do seu corolário: manejo correto de métodos e domínio das noções de inferência e causalidade. Soares (2005) alertara para o descaso e o despreparo que marcavam as Ciências Sociais brasileiras no que diz respeito a métodos. Desde esse primeiro diagnóstico, houve um progresso considerável no que tange à disseminação de cursos de métodos nas universidades (BARBOZA & GODOY, 2014)¹. Contudo, mesmo passada uma década, ainda são minoritárias as publicações em Ciência Política e Relações Internacionais metodologicamente rigorosas (NEIVA, 2015; MEDEIROS et al 2016).² Dessa forma, continua sendo uma tarefa importante capacitar os recém-ingressos neste campo de atividade científica no manuseio adequado dos seus pressupostos e ferramental.

Neste espírito, o presente trabalho se propõe a realizar um resumo didático e introdutório das principais discussões da literatura sobre métodos em Ciência Política e Relações Internacionais. Em especial, revisamos as diretrizes propostas pelos autores canônicos do campo³, bem como os aperfeiçoamentos e contra-argumentos desenvolvidos por outros pesquisadores acerca dessas orientações. Inspira-se, portanto, fortemente na discussão iniciada por King, Keohane e Verba (1994) em *Designing Social Inquiry* e nas suas repercussões. Passados mais de 20 anos de sua primeira edição, o livro teve um forte impacto no campo, moldando a agenda de pesquisa da disciplina em termos preocupação com desenhos de pesquisa e inferência (REZENDE, 2017).

Buscaremos, ao longo do artigo, explicar os conceitos de (1) *desenho de pesquisa*, (2) *inferência*, e (3) *causalidade*, sendo o texto também estruturado nestes três tópicos.

1 O levantamento de Barboza e Godoy (2014) mostra que, entre 1998 e 2012, instituições como a UFMG, UFPE e IUPERJ expandiram sua oferta de disciplinas metodológicas. Neiva (2015) também verifica que autores afiliados a essas instituições, juntamente com a UFPR, tenderam a publicar mais artigos quantitativos entre 1997 e 2012.

2 Neiva (2015) detecta que o uso da estatística avançada em artigos de ciências sociais no Brasil vem crescendo (de 3,3% em 1997 para 7,6% em 2012), porém o campo segue majoritariamente “não quantitativo”. Medeiros et al (2016) mostraram que menos de 13% dos artigos em Relações Internacionais na América do Sul mencionam algum método explicitamente.

3 Convém destacar que os principais autores trabalhados possuem uma perspectiva positivista sobre o saber científico, de modo que está fora do escopo dessa revisão as contribuições feitas por autores pós-positivistas, que problematizam a possibilidade de conhecimento científico objetivo, dissociado de processos históricos, relações de poder e ideologia (ALMOND, 1988).

1 DESENHO DE PESQUISA

Rezende (2017, p. 248) define desenho de pesquisa como “estratégias de identificação onde são estabelecidas premissas que permitem identificar efeitos causais específicos e contém as provas formais que tornam possível fazer com que estes conduzam ao efeito causal em amostras infinitas”. Essa conceituação oferece-nos um ponto de partida, ressaltando as linhas mestras de um desenho de pesquisa: identificação, causalidade e elaboração formal⁴.

Para compreender como desenhos de pesquisa levam à produção de conhecimento científico, veremos nesta seção (1) os critérios que conferem cientificidade a uma pesquisa em Ciência Política e Relações Internacionais; (2) os componentes que devem estar presentes no desenho de pesquisa; (3) em torno de quais aspectos desenhos de pesquisa tendem a divergir e estabelecer tradições separadas; e (4) alguns problemas comuns de desenhos de pesquisa.

1.1 O que torna uma pesquisa científica?

O que torna a investigação científica distinta de outras formas de conhecimento? A indagação tem uma longa tradição na filosofia da ciência. Há mais de um século, escreveu Karl Pearson (1900) que a unidade de toda a ciência reside em seu método, não no seu material. Em 1942, Robert Merton postulou que o conhecimento científico deve ser definido pelo seu caráter comunal, universalismo, impessoalidade e por estar sujeito ao “ceticismo organizado”.

Em *Designing Social Inquiry*, King, Keohane e Verba (1994) — doravante “KKV” — definem o que é a pesquisa científica a partir de quatro critérios: (1) seu objetivo é chegar à inferência causal ou descritiva; (2) seus procedimentos são públicos, e é a transparência que permitirá à comunidade científica julgar sua validade, aprender do método, replicar seus resultados, compreender suas limitações para tentar superá-las, e fazer comparações com diferentes pesquisadores⁵; (3) as conclusões são incertas, visto que a inferência a partir de dados imperfeitos não pode levar a conclusões perfeitas; e (4) seu conteúdo é o método, ou seja, a cientificidade de uma pesquisa está nos métodos e regras empregados, não no seu assunto, e é por isso que se pode chamar de ciência uma prática tão multifacetada e que contempla tantas de áreas do saber.

1.2 O que um desenho de pesquisa precisa ter?

KKV (1994) definem o desenho de pesquisa como composto por quatro elementos. O primeiro elemento é a *pergunta de pesquisa*, e esta deve ser relevante tanto para a sociedade quanto para a academia. Isto é, seus achados devem impactar a vida das pessoas de forma efetiva tratando de um tema que importe. Ademais, a pergunta deve contribuir para o estado da arte científico. Precisa, portanto, estar em diálogo com a literatura existente e buscar avançar o conhecimento disponível. Assim, o pesquisador deve precaver-se, por um lado, contra querer tratar um tema pertinente (e.g.: guerras, fome, pobreza), porém sem a sistematicidade que torna a investigação propriamente científica, e, por outro, contra se consagrar à solução de quebra-cabeças catedráticos de pouca importância fora dos muros da universidade.

O segundo elemento é a *teoria*, definida por KKV como uma especulação precisa e racional sobre a resposta a uma questão de pesquisa, e que inclui uma declaração sobre por que a resposta proposta é correta. São propriamente

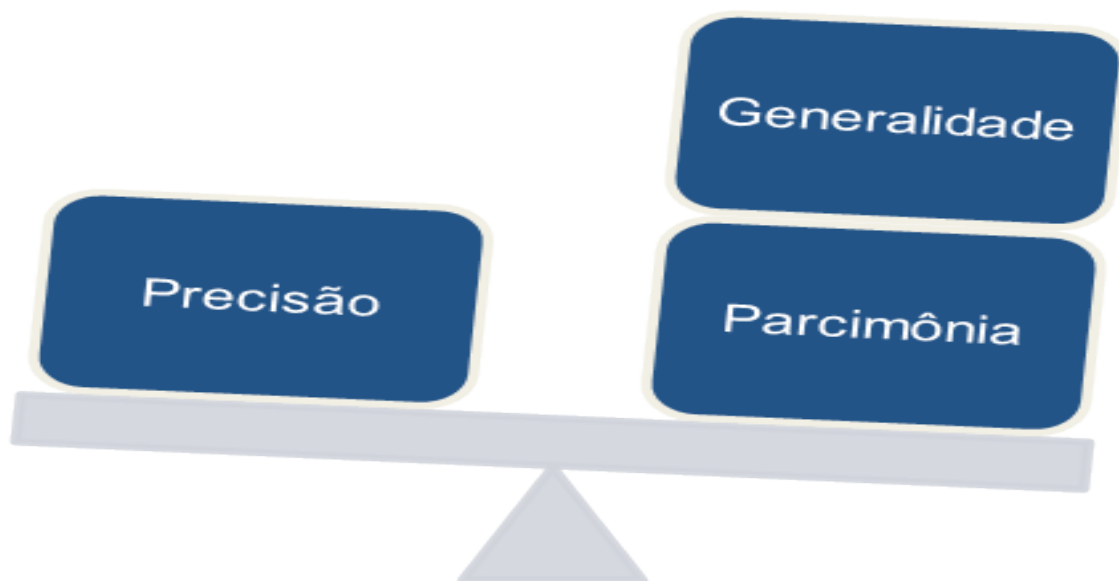
4 Convém, não obstante, ressaltar que o desenho de pesquisa não deve ser confinado apenas ao âmbito causal: estudos voltados à inferência descritiva também necessitam mobilizá-lo em seu esforço de identificação.

5 King (2015) [1995] argumentou posteriormente sobre o imperativo de se instaurar a replicação de resultados como prática padrão da Ciência Política. A *Revista Eletrônica de Ciência Política* traduziu o artigo original publicado na *PS: Political Science & Politics* em um número especial sobre Metodologia de Pesquisa (vide referências). Também em 2015, a *Revista Política Hoje* editou um dossiê especial sobre o assunto (<https://periodicos.ufpe.br/revistas/politica hoje/issue/view/722/showToc>).

científicas as teorias *falsificáveis*. Tal orientação remota a Karl Popper, que em 1935 argumentou que sistemas de conhecimento precisam poder ser falsificados a partir da experiência, para que sejam considerados empíricos e científicos. Assim, o método empírico cumpre seu papel quando expõe um sistema à possibilidade de ser contraditado. Portanto, recomenda-se privilegiar teorias que gerem o máximo possível de implicações observáveis. Desse modo, pode-se potencializar as oportunidades de verificação dos desdobramentos empíricos dessas conjecturas teóricas e assim de corroborá-la ou invalidá-la.

Przeworski e Teune (1970) também buscaram definir bons padrões para avaliar teorias em pesquisa comparada, reduzindo-os a quatro objetivos cardinais: (1) precisão (o grau de detalhamento atingido na explicação dos casos); (2) generalidade (o grau de abrangência da explicação); (3) parcimônia (conseguir explicar um número alto de casos a partir de um conjunto pequeno de causas) e (4) causalidade (estabelecer uma relação de causa e efeito entre as variáveis mobilizadas). Não é possível perseguir concomitantemente todos os objetivos com o mesmo êxito. Pesquisas que buscam ser parcimoniosas e gerais não serão igualmente precisas, pois mobilizar poucas variáveis permite abranger um número alto de casos, porém, quanto menos fatores e mais casos, menor será o nível de detalhamento possível. Alternativamente, maior precisão requer maior detalhamento de casos cada vez mais particulares, implicando, portanto, menor generalização e parcimônia. Há um *trade off* entre esses objetivos e cabe ao pesquisador eleger o que é mais relevante para o seu intento. A Figura 1 ilustra a relação entre precisão e generalidade.

Figura 1: *Trade off* entre Precisão vs. Generalidade e Parcimônia



Fonte: Elaborado pelo autor, com base em Przeworski e Teune (1970)

O terceiro componente apontado por KKV são os *dados* coletados empiricamente para teste das hipóteses previstas na teoria. É importante destacar que a etapa incontornável entre a teoria e os dados é a *mensuração*. Esta é para Miller (2011) o coração do desenho de pesquisa, pois permitirá unir o teórico e o empírico.

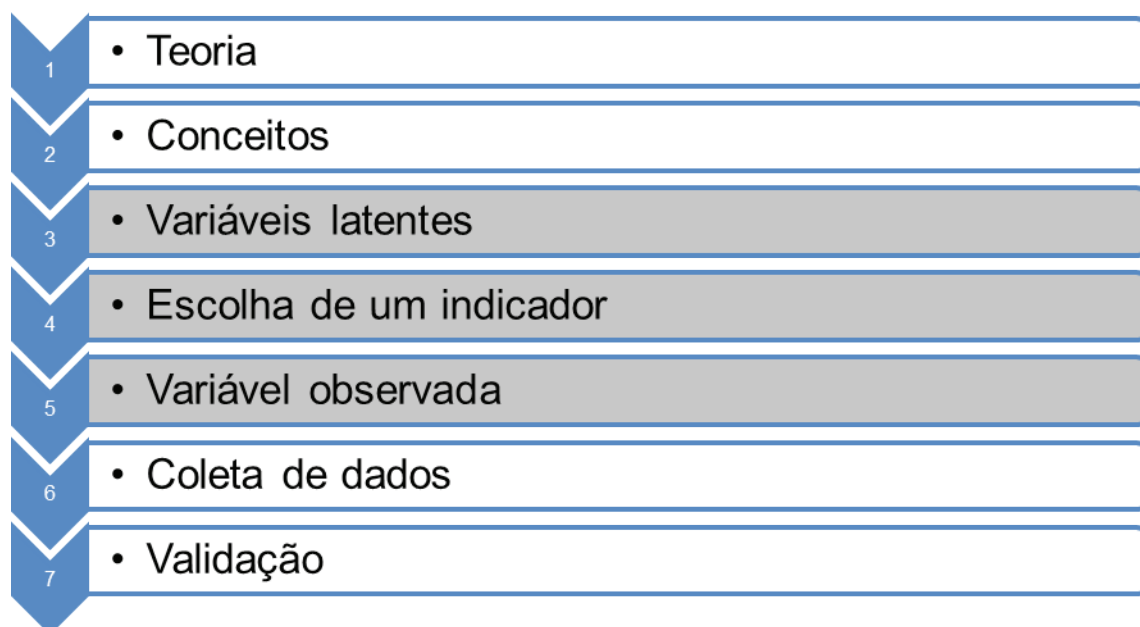
O que é, portanto, medir? Miller propõe uma analogia do mundo natural: suponhamos que há um conceito abstrato que se queira mensurar (temperatura), porém ele é inobservável diretamente. Sabendo-se que o volume do mercúrio é afetado pela temperatura (implicação observável), usa-se aquele para medir esta. A opção pelo mercúrio é chamada de *operacionalização*. Feita a medição, é preciso verificar se seu resultado corresponde realmente à mudança da

temperatura, o que é o problema da *validação*, ou seja, a medida registrada reflete o conceito que se tinha em mente. Por fim, é preciso garantir que os mesmos resultados serão obtidos por outros pesquisadores nas mesmas condições. A isto se chama *confiabilidade*.

Passando da natureza à política, encontramos o mesmo movimento, partindo de conceitos abstratos a implicações acessíveis. Dahl (1957), ao tratar do “poder” – conceito tão nebuloso quanto onipresente na Ciência Política – inicia definindo-o da seguinte forma: A tem poder sobre B se consegue levar B a fazer algo que B em outro caso não faria. Tem-se aí uma conceituação que produz uma implicação observável (a mudança de comportamento em B) e pode ser, portanto, operacionalizada (e.g.: observar como vota um parlamentar B depois de interagir com A, líder da sua bancada). Em Relações Internacionais, Keohane e Nye (2001) afirmam que as relações interestatais hoje são caracterizadas pela interdependência e esta se observa em dois componentes: sensibilidade (exposição a efeitos custosos de mudanças geradas no exterior) e vulnerabilidade (custo para adaptar-se face às mudanças geradas no exterior). A sensibilidade e a vulnerabilidade de um país, por exemplo, no campo de recursos energéticos, podem então ser operacionalizadas observando, respectivamente, quanto do petróleo consumido em um país é importado e qual a viabilidade de implementar políticas que aumentem ou diminuam essa proporção.

Organizando em etapas, o cientista social precisa, portanto, (1) partir de uma teoria para (2) desenvolver conceitos claros e precisos sobre o que se quer medir. Desses conceitos serão derivadas (3) variáveis latentes, que são aquilo que o observador idealmente gostaria de observar⁶. Como frequentemente estudamos fenômenos que não são diretamente observáveis, recorre-se a um (4) indicador que leva à (5) variável observada. O processo de operacionalização vai das etapas 3 a 5. Após a (6) coleta de dados, é preciso assegurar-se que (7) eles são válidos e confiáveis. Esse procedimento está resumido na Figura 2.

Figura 2: Etapas do processo de mensuração



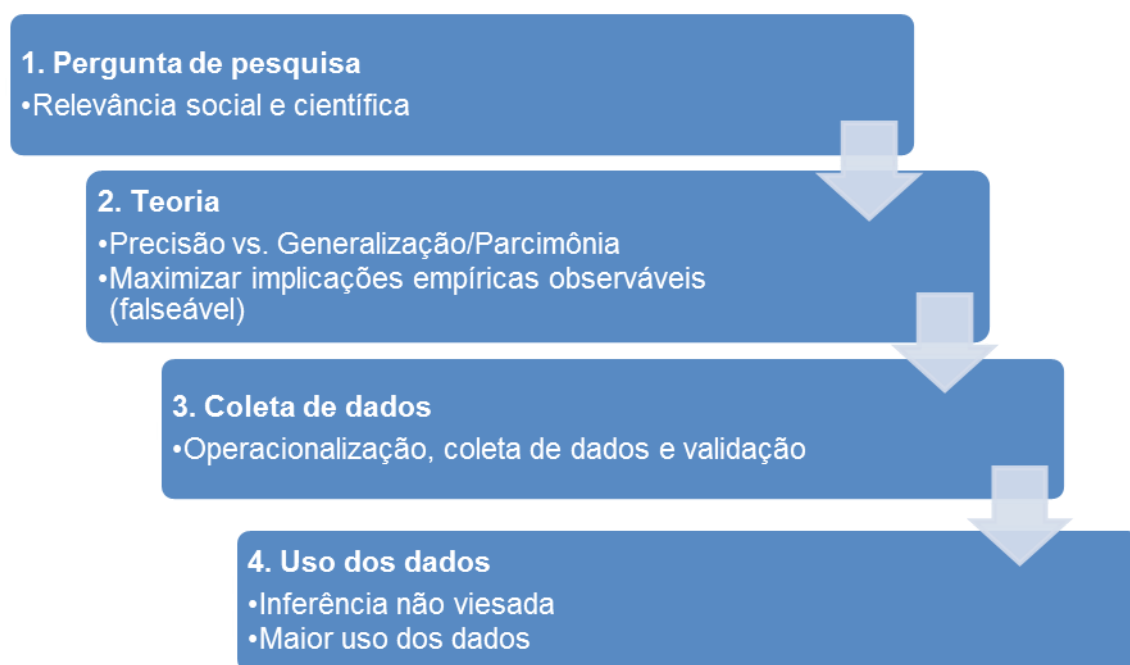
Fonte: Elaborado pelo autor, baseado em Miller (2011). Células em cinza representam processo de operacionalização.

⁶ Miller (2011) sustenta que variáveis latentes não são sinônimas de conceitos, pois há conceitos que requerem mais de uma variável latente para serem mensurados.

É preciso, entretanto, ter em mente que validade e confiabilidade são particularmente fugidias fora de um contexto experimental – como é o caso das Ciências Sociais. Isso pois o pesquisador, desejando encontrar no mundo empírico determinado conceito (e.g.: “discriminação racial”), coleta valores de algum indicador (“renda”), mas estes valores são produzidos não somente pelo conceito de interesse, mas por uma miríade de *confounders* (“educação”, “habilidade”, etc). Por esse motivo, Blalock (1961, p.347) afirma que uma “medida satisfatória de um conceito teoricamente definido pede uma teoria de como outras variáveis estão operando”⁷.

Por fim, o quarto e último elemento do desenho de pesquisa apontado por KKV é o *uso dos dados*. O melhor emprego dos dados ocorre quando se busca fazer inferências não viesadas e maximizando o uso das informações coletadas – questões que serão melhor detalhadas nas seções sobre inferência e causalidade. Em síntese, as etapas descritas pelos autores podem ser sintetizadas da seguinte forma:

Figura 3: Elementos do desenho de pesquisa



Fonte: Elaborado pelo autor, baseado em Przeworski e Teune (1970), KKV (1994) e Miller (2011)

1.3 Diferenças entre desenhos de pesquisa

Há muitas perguntas em Ciência Política e Relações Internacionais, assim como há muitas maneiras de respondê-las. Assim, os desenhos de pesquisa disponíveis aos pesquisadores também são diversos, cada qual possuindo suas vantagens e limitações. Dentre os muitos quesitos a partir do qual se podem diferenciar desenhos de pesquisa, enfocaremos neste trabalho (a) método, (b) número de observações e (c) o uso ou não de experimentos, já que estes acarretam consequências importantes em termos de justificativa do desenho utilizado, validade inferencial e causalidade.

⁷ Tradução do autor a partir do original: “a satisfactory measure of a theoretically defined concept calls for a theory of how other variables are operating” (BLALOCK, 1961, p.347)

a) Método: Qualitativo x Quantitativo

A divisão entre métodos qualitativos e quantitativos é um longo tema de debate nas Ciências Sociais e foi uma das motivações para a escrita de *Designing Social Inquiry*⁸. A obra asseverava que, na visão dos autores, ambos os métodos diferem apenas em estilo e partilham de uma mesma lógica inferencial; e se destinava a prover subsídios para que pesquisadores quali (que, segundo os autores, tendem a ser menos rigorosos metodologicamente) fossem mais ciosos de seus desenhos de pesquisa e assim aprendessem como maximizar a validade dos seus achados (KKV, 1994). Muitas das críticas dirigidas ao projeto de KKV, condensadas notadamente no *Rethinking Social Inquiry* de Brady e Collier (2004), alegaram que a lista sugerida de boas práticas se baseava em um referencial quantitativo, voltado à regressão linear, e que assim negligenciava virtudes próprias às pesquisas quali, por exemplo o potencial dos estudos de caso para a generalização e a elaboração de hipóteses *post hoc*. Alguns dos autores do volume editado por Brady e Collier compartilham com KKV a visão de uma lógica inferencial única e preocupam-se em corrigir e aprimorar as recomendações para aliar ambas as tradições. Tarrow (2004), por exemplo, lista um ferramental de estratégias que podem ser utilizadas para combinar métodos quali e quanti. Outros autores, porém, entendem que o fosso entre ambos os lados é maior, e insistiram numa especificidade da lógica qualitativa que não permitiria o transporte de conceitos de um campo para o outro. Essa excepcionalização se pautava, primeiramente, com respeito ao número adequado de observações para propor relações causais e, em segundo, em considerações ontológicas sobre a natureza da realidade: se determinista ou probabilística⁹.

b) Número de observações: Large n x Small n

Pesquisas *large n* são (quase) exclusivamente quantitativas, pois métodos estatísticos são mais viáveis para abranger observações abundantes. Por contemplar um número grande de casos, naturalmente sacrificam em detalhamento e profundidade o que ganham em generalidade e, mais importante, em alcance inferencial sobre a validade da relação estudada. Portanto, as explicações produzidas por tais estudos não possuem grande especificidade, ou seja, não explicam exaustivamente as causas envolvidas na produção de um fenômeno específico, sendo antes mais adequadas para entender o comportamento de uma relação causal numa população. São, portanto, mais aptos para mensurar os *efeitos das causas* (i.e.: “o quanto X afeta Y?” ou “o quanto de Y pode ser explicado por X?”), sendo também chamados de orientados a variáveis ou orientados à população (MAHONEY, 2008; RAGIN, 2004).

A forma consagrada de estudos quantitativos *large n* é a regressão de mínimos quadrados ordinários (MQO), que sozinha está presente em quase um terço da produção em Ciência Política nos Estados Unidos (KRUEGER & LEWIS-BECK, 2008 *apud* FIGUEIREDO FILHO et al, 2011, p. 46). A análise de regressão consiste na investigação da relação potencial existente entre variáveis. Mais especificamente, ela permite analisar a associação entre uma variável dependente e uma ou mais variáveis independentes, sua magnitude (se forte ou fraca) e direção (se negativa ou positiva). A utilidade da análise de regressão reside, primeiramente, em sua simplicidade para fins de interpretação, em seguida, na sua robustez à violação de alguns pressupostos, e também em permitir ao mesmo tempo explicar, interpretar e prever valores para as variáveis de interesse (FIGUEIREDO FILHO et al. 2011, p.48-49)¹⁰.

Já pesquisas *small n* possuem um número reduzido de casos. Elas lançam mão principalmente de métodos qualitativos e ganham em profundidade e conhecimento do contexto o que perdem em generalização e parcimônia. Devido à sua

8 Ver também Almond (1988) com respeito a divisões metodológicas e ideológicas da Ciência Política norte-americana.

9 Vide também Silva (2015) para uma discussão das dificuldades dos métodos mistos, relativas à (i) posição filosófica adotada com respeito à relação entre metodologia e expectativas de verdade; e ao (ii) pluralismo metodológico do próprio campo qualitativo.

10 Ver também Cap. 2 de Wooldridge (2013) para uma introdução prática

atenção intensiva a poucos casos, é um tipo de pesquisa que atenta mais para a produção de resultados específicos. Em outras palavras, preocupa-se com as *causas dos efeitos* (i.e.: “o que causa Y?”), sendo por isso também chamadas de orientadas a casos (MAHONEY, 2008; RAGIN, 2004).

Em se tratando de número de observações, merecem atenção especial os estudos de caso. Estes são tradicionalmente classificados como estudos *small n* em que $n = 1$ (LIJPHART, 1971). Gerring (2004) propõe redefini-los como um estudo intensivo de uma unidade com o objetivo de se compreender uma classe maior de unidades similares. Como unidades são por ele definidas como podendo conter vários casos, o autor concebe que estudos de caso podem ser tanto *large n* quanto *small n*. Sua lógica é detectar covariação, e esta pode ser buscada no interior de uma única unidade (*within unit*), comparando diferentes unidades (*across units*), combinando ambos, estaticamente ou ao longo do tempo para todas as unidades, produzindo assim diferentes tipos de estudo.

Para Lijphart (1971), os estudos de caso que mais acrescentam à teoria são aqueles que buscam, a partir das observações em uma unidade, elaborar postulados que possam ser estendidos a mais unidades (“estudos geradores de hipóteses”), e aqueles que investigam casos aberrantes, que se afastam do padrão apontado pela teoria e, por conseguinte, podem encerrar informações importantes sobre a relação causal em questão (“estudo de casos desviantes”). Gerring (2004) acrescenta que estudos de casos serão tanto mais relevantes quanto mais restrito for o número de casos empíricos que demonstrem variabilidade útil para a variável de interesse.

c) Experimento x Observação

Estudos experimentais buscam descobrir se X causa Y recorrendo à manipulação deliberada do pesquisador: submete-se um grupo à exposição da causa potencial e outro não (grupos “tratamento” e “controle”, respectivamente). A diferença média nos valores da variável dependente entre os grupos será chamada de causa. A principal vantagem deste desenho está no fato que o cientista determina a atribuição de valores na variável independente e, especialmente, a seleção das observações, o que lhe permite recorrer à randomização como forma de assegurar que a regra de seleção não estará correlacionada de nenhuma forma com qualquer variável, livrando-se assim de vários vieses.

Já em estudos observacionais, o pesquisador não possui o poder para manipular o contexto. Antes, defronta-se com um universo já moldado pela história e/ou pela natureza. Consequentemente, deve lançar mão de um outro arsenal de técnicas para conseguir estabelecer causalidade, observando no mundo empírico as diversas manifestações disponíveis da relação entre X e Y e assim buscando tecer inferências sobre sua interação. Enquanto que para as Ciências Naturais o uso de experimentos é a norma, estes são raros para a Ciência Política e Relações Internacionais por motivos tanto práticos quanto éticos¹¹.

1.4 Alguns problemas de desenhos de pesquisa: viés de seleção e viés de variável omitida

Um problema recorrente, cujo tratamento adequado é alvo de contenda na comunidade científica, é o viés de seleção. Este ocorre quando a seleção dos casos é afetada pela variável dependente, ou seja, estudam-se apenas casos em que o fenômeno de interesse se manifesta. Como demonstrado por Geddes (1990), o resultado pode ser uma subestimação do efeito causal. Embora pareça intuitivamente produtivo estudar só os casos positivos (e.g.: se desejo entender o que

11 Não obstante, tem se popularizado a aplicação de desenhos de pesquisa “semi-experimentais”. De acordo com dados do *Google Scholar*, apenas 338 artigos continham os termos “*quasi-experiment*” e “*political science*” em 2005. Em 2015, eram 1520 (pesquisa feita pelo autor em scholar.google.com.br). Estes também são chamados “quase experimentais” ou “experimentos naturais” segundo o caso e tipicamente não permitem ao pesquisador garantir a aleatoriedade da amostra ou ter controle sobre a aplicação do tratamento (e.x.: comparar os efeitos no número de acidentes na estrada antes e depois após introdução de nova lei de trânsito). Concomitantemente, expedientes estatísticos como *Difference in Difference*, *Propensity Score Matching*, regressão descontinuada e controle sintético também buscam verificar o efeito causal de tratamentos.

explica vitórias eleitorais, por que incluir no estudo candidatos perdedores?), isso reduz a variabilidade da variável dependente (COLLIER & MAHONEY, 1996). A seleção baseada na variável independente, alternativamente, não enviesaria os resultados. Embora o viés de seleção seja geralmente fruto de escolhas conscientes ou inconscientes do pesquisador, há vieses que são um produto da ação da história ou da natureza, que privilegiaram certas manifestações de um dado fenômeno e suprimiram outras.

Para alguns autores ligados à tradição de estudos de caso, entretanto, esses alertas precisam ser revistos. Brady e Collier (2004) defendem que as advertências feitas por KKV contra vieses de seleção, embora úteis, se aplicam especificamente às pesquisas qualitativas entre casos distintos (*cross-case*). Já para os estudos aprofundados em um caso singular (*within-case*), a truncagem da amostra não obscureceria, necessariamente, a relação entre as variáveis independente e dependente, pois este tipo de análise se vale de observações de processos causais, que aderem a uma lógica alternativa à regressão linear. Não obstante, amostras truncadas podem superestimar certos casos, de modo que continua sendo imprescindível aos pesquisadores quali serem zelosos na escolha de suas observações. Pela mesma lógica, os autores defendem a validade dos estudos *within-case* mesmo quando padecem do problema de nenhuma variância na variável dependente, já que isto é um problema maior para as pesquisas que culminam em alguma forma de regressão.

A seleção de casos com base no valor das variáveis é uma questão importante, quer os estudos tenham uma abordagem probabilística, clássica dos estudos quantitativos orientados à população, ou de causas suficientes e necessárias, comum em estudos qualitativos orientados a variáveis. Selecionar casos a partir da variável dependente leva à identificação de causas necessárias — pois sempre que ocorrer uma, ocorrerá a outra —, mas para se saber sobre causas suficientes é preciso admitir variabilidade na variável dependente. Pérez-Liñán (2010) descreve como esse procedimento, em que se sucedem os métodos de semelhança e de diferença de J. S. Mill, pode ser complementado com a Análise Qualitativa Comparada (QCA) para chegar a um número de configurações causais que explicam a ocorrência de um fenômeno¹².

Um segundo viés problemático é o das variáveis omitidas, pois elas podem invalidar as relações causais propostas por um desenho de pesquisa. Se se postula que determinado X causa Y, mas o pesquisador falha em perceber que há uma outra variável Z intervindo nesta relação, por exemplo sobredeterminando os valores de X, ou impactando simultaneamente X e Y, a relação apontada será na verdade espúria. Em suma, se o fenômeno estudado é melhor descrito por um conjunto de variáveis e alguma dessas é deixada de fora da equação, o efeito das variáveis preservadas no modelo será ou super ou subestimado (WOOLDRIDGE, 2013, p. 88).

2 INFERÊNCIA

A inferência pode ser entendida, resumidamente, como chegar a conclusões sobre fenômenos não observados a partir de informações sobre fenômenos observados (KKV, 1994). Ao realizar a inferência, o pesquisador caminha na direção da generalização de seus achados, uma vez que tenta alargar os padrões vistos num caso particular para o geral. Assim definida, constata-se que a inferência é um objetivo central da pesquisa científica para diferentes metodólogos. Veremos nesta seção como alguns desses autores (1) justificam o movimento em direção à generalização e lidam, neste percurso, com o problema de garantir a comparabilidade entre casos; e em seguida analisaremos (2) os dois tipos de inferência previstos por KKV, e (3) recomendações da literatura sobre como melhorar inferências.

2.1 Particularizar x Generalizar

Przeworski e Teune (1970, p.4) defendem que o principal propósito da pesquisa social é “levar a afirmações gerais sobre fenômenos sociais”. Consequentemente, se aceita que o *explanandum* das Ciências Sociais, o comportamento

12 Vide também Freitas e Neto (2016) para uma apresentação didática do método.

humano ou social, pode ser explicado a partir de um número discreto de *explanans*: leis gerais estabelecidas pela observação.

A busca de leis generalizáveis é uma meta própria das Ciências Naturais, por exemplo a Física, que procura padrões e leis causais (*covering laws*) que se manifestem a despeito das contingências situacionais (ALMOND & GENCO, 1977). Isto é, a presença das causas deve conduzir sempre ao resultado previsto, com nenhuma ou quase nenhuma variação não explicada. Transpor esse projeto para as Ciências Sociais não é sem controvérsia, uma vez que estas se originaram da Filosofia — em especial da Filosofia da História — e, portanto, herdaram maior preocupação precisamente com as excepcionalidades e particularidades, ao invés dos universais. Em termos técnicos, o debate que perpassa as Ciências Sociais é se sua vocação é nomotética, voltada à proposição de leis gerais, ou idiográfica, concentrada nas contingências¹³.

Almond e Genco (1977) recuperam uma metáfora popperiana que captura bem a dicotomia. Há fenômenos que se comportam regulares como relógios, de forma determinada e previsível, e outros caoticamente como nuvens, indeterminados e com pouca previsibilidade. A preocupação com a descoberta de leis gerais leva o cientista social a preferir “relógios” a “nuvens”, o que, na avaliação de Almond e Genco, é problemático, já que as interações humanas e sistemas sociais raramente terão a invariabilidade da mecânica e, diferente de moléculas em um laboratório, serão movidas por fatores como vontade, finalidades, estratégias e crenças. Daí a importância de se pensar o mundo social relaxando expectativas e assim acomodando irregularidade e intencionalidade.

Przeworski e Teune (1970) reconhecem que, nas Ciências Sociais, proposições que tentem explicar um fenômeno de forma universal são julgadas descabidas. Sua atenção, contudo, não recai sobre a questão da irregularidade, mas principalmente da diversidade. Ou seja, regras generalizantes podem ser indevidas por ignorarem a multiplicidade de formas em que certo fenômeno se manifesta em diferentes sistemas sociais. Por exemplo, certo comportamento pode ser classificado em um país como “corrupção”, mas ser aceitável em outro.

As categorias mobilizadas para propor uma lei, observadas em determinado contexto, não são automaticamente transladáveis a outros cenários. Subjaz essa previsão uma crença que sistemas sociais são diversos entre si ao ponto que, ainda que se dê um único nome a certo fenômeno de interesse (e.g.: “protesto”, “democracia”), sua materialização em ambientes distintos mobilizará tantas idiossincrasias locais que o conceito já não servirá para descrever a mesma coisa.

Com respeito a essa querela, Przeworski e Teune (1970) assumem a posição que a pesquisa comparada pode sim produzir afirmações nomotéticas. Em outras palavras, ao invés de se contentar em afirmar que certos fenômenos se observam em sistemas específicos, isto é, utilizar os nomes próprios dos sistemas (ex.: democracia na Índia, Canadá, Nigéria, etc) como previsores, é possível substituir estes nomes por variáveis mensuráveis.

Buscar a generalização não é prescindir da história em prol do abstrato. Antes, enunciados historicamente situados são potencialmente generalizáveis, pois encerram sob nomes próprios (ex.: “Alemanha no entre-guerras”, “África nos anos 1960”, etc.) um conjunto de fatores causais latentes. Estes podem ser transportados de um contexto a outro de modo a se alcançar uma teoria mais geral, restando os fatores únicos e irreplicáveis como um “resíduo” da explicação teórica. Por isso, os Przeworski e Teune¹⁴ defendem que o elo entre observações históricas e teorias gerais está na substituição dos nomes próprios por variáveis relevantes (quando se acredita que dada classe de eventos sociais pode ser expandida para além de um sistema específico), e que este é o fim da pesquisa comparada. Para eles, o êxito em comparar um mesmo elemento em situações diferentes depende dos critérios de confiabilidade e validade, já que para

13 Przeworski & Teune, *idem*

14 *Idem*

substituir nomes próprios de sistemas sociais por variáveis analíticas é preciso assegurar-se que estas de fato capturam o fenômeno de interesse.

Pode-se dizer que duas coisas serão comparáveis com respeito aos atributos que elas compartilham: se não compartilham nenhum, são incomparáveis, se todos, são a mesma coisa, e em ambos os casos a comparação é inútil. Para as situações entre esses extremos, ela é viável. Sartori (1991), contudo, alerta para o risco de o pesquisador, na busca por aumentar o número de casos sob análise, estique um conceito ("*concept stretching*") ao ponto de esvaziá-lo de qualquer precisão semântica, tornando-o operacionalmente inválido¹⁵.

2.2 Inferências descritivas e inferências causais

KKV (1994) afirmam que o objetivo da pesquisa científica é a inferência com base em informações empíricas. Os dados acumulados são subsídio para inferências sobre algo além do que é diretamente visto.

O trio de autores reconhece que há dois tipos de inferência. O primeiro é a inferência descritiva, na qual se usam observações do mundo para aprender sobre fatos não observados. Descrever é também inferir se considerarmos que a categorização e tipificação desenvolvidas com base em um número restrito de fenômenos poderá, possivelmente, contemplar outros fenômenos ainda não observados.

O segundo tipo de inferência prevista por KKV é a causal. Na opinião dos autores, esta é desejável sempre que possível no intuito de complementar o quadro inicialmente oferecido pela descrição. A descrição em KKV é, portanto, uma etapa anterior à explicação causal.

Os autores propõem três critérios para julgar a validade das inferências. Primeiramente, as inferências devem ser *não-viesadas*. Há viés quando a replicação produz erro sistemático que, no limite, afastará o resultado da média original¹⁶. Segundo, para diferenciar entre estimadores igualmente livres de viés, convém selecionar com base na *eficiência*, ou seja, quando há baixa variância. A variância será menor nos estudos com muitas observações, o que é uma consequência do terceiro critério, a *consistência*, que prevê que quanto maior o número de observações menor a variabilidade.

2.3 Melhorando inferências: indeterminação, número de observações e multicolinearidade

KKV (1994) advertem sobre o problema de desenhos de pesquisa indeterminados. Estes ocorrem em dois casos: quando se tenta realizar mais inferências do que o número disponível de observações permite e quando há multicolinearidade.

O primeiro problema diz respeito a um dilema apontado por Lijphart (1971, p.686): "muitas variáveis, N pequeno". Quando se tem poucas observações, é difícil sustentar que a relação causal proposta para os casos se explica por um motivo e não outro. Assim, a determinação de um desenho de pesquisa diz respeito à proporção entre variáveis e observações. Por isso, KKV e outros autores recomendam aumentar o número de observações como forma de resolver o problema da indeterminação¹⁷. KKV sugerem que, quando aumentar o N não for possível, deve-se mudar a ênfase da pesquisa para que, ao invés de descobrir as causas dos efeitos, se tente mensurar os efeitos das causas, ou ainda que se busquem mais observações em outros níveis de análise (por exemplo, dados em nível estadual ao invés de nacional).

15 Collier e Mahon (1993) revisitaram o problema do alargamento conceitual e propuseram alguns construtos adicionais para permitir a comparabilidade entre fenômenos, como os conceitos de "semelhança familiar" e "categorias radiais".

16 Mais especificamente, o viés ocorre quando as amostras coletadas de uma população geram estimadores que estão sistematicamente afastados dos valores reais dos parâmetros, tal qual distribuídos na população (KENNEDY, 2003).

17 Vide também Sartori (1991).

Repetir observações ao longo do tempo (criando os chamados painéis de dados ou *Time-Series Cross-Section Data*) é outra estratégia para multiplicar o N.

Munck (2004) problematiza essa solução afirmando que ela nem sempre é viável: há fenômenos políticos que contabilizam poucos exemplares na história humana. Revoluções e guerras, por exemplo, ocorrem apenas raramente. Ademais, tentar colocar o maior número de casos debaixo de um mesmo rótulo pode também aumentar os riscos de se incorrer no alargamento de conceitos ou contextos de heterogeneidade causal, isto é, situações em que o fenômeno passa a ser produzido por causas distintas. A preocupação subjacente de KKV é buscar maior alcance inferencial; porém, argumenta Munck (2004), para pesquisas quali isso adviria mais pelo conhecimento intensivo do pesquisador sobre o caso e contexto, do que pela inclusão de mais casos.

Collier, Brady e Seawright (2004), por sua vez, também problematizam a inclusão de mais casos como forma de melhorar a inferência, e privilegiam a noção de que o conhecimento contextual é igualmente importante na validade inferencial. Eles propõem, portanto, uma distinção entre observações de banco de dados (*data-set observations* - DSO) e observações de processos causais (*causal-process observations* - CPO). As primeiras são as observações no sentido tradicional estatístico (i.e.: pontos num gráfico de dispersão) e fornecem a base para as inferências causais correlacionais¹⁸. Já as CPO são definidas como *insights* ou informações que instruem sobre o contexto e mecanismos, também contribuindo para a compreensão do processo causal, embora não venham a ser tabuladas como novos casos ou valores de uma variável explicativa. Enquanto que aumentar o número de DSOs atende o requisito clássico de aumentar o N, elevando por conseguinte os graus de liberdade, aumentar as CPOs não afeta o N nem os graus de liberdade¹⁹; assim, defendem os autores, ambos os procedimentos subsidiam inferência, embora por lógicas distintas.

A segunda ameaça à inferência é a multicolinearidade. Há multicolinearidade quando há alto nível de correlação entre as variáveis independentes²⁰. Sua presença pode trazer sérios problemas para o alcance inferencial do modelo explicativo mobilizado pelo pesquisador.

A maior dificuldade de modelos com problemas de multicolinearidade é o aumento da magnitude da variância dos parâmetros estimados. Isso porque a presença de altos níveis de correlação entre as variáveis independentes impossibilita estimar, com precisão, o efeito de cada variável sobre a variável dependente [...] quanto maior a correlação entre as variáveis independentes, menos informação estará disponível para estimar os coeficientes associados às variáveis explicativas. (FIGUEIREDO FILHO et al., 2011, p.57-58)

Segundo KKV (1994), a seleção aleatória das observações é um dos mecanismos na tradição quantitativa para evitar indeterminações dessa natureza, porém os autores reconhecem que para alguns fenômenos analisados pela tradição qualitativa tal procedimento pode ser pouco prático. Assim, eles sugerem neste caso coletar mais observações (preferencialmente casos em que as variáveis explicativas não estejam fortemente correlacionadas), seja no mesmo nível de análise ou em outro. Alternativamente, pode-se limitar o número de variáveis utilizadas – pelo menos até onde for

18 Ferramenta pedagógica para visualizar correlação e *data points*: <<http://guessthecorrelation.com/>>

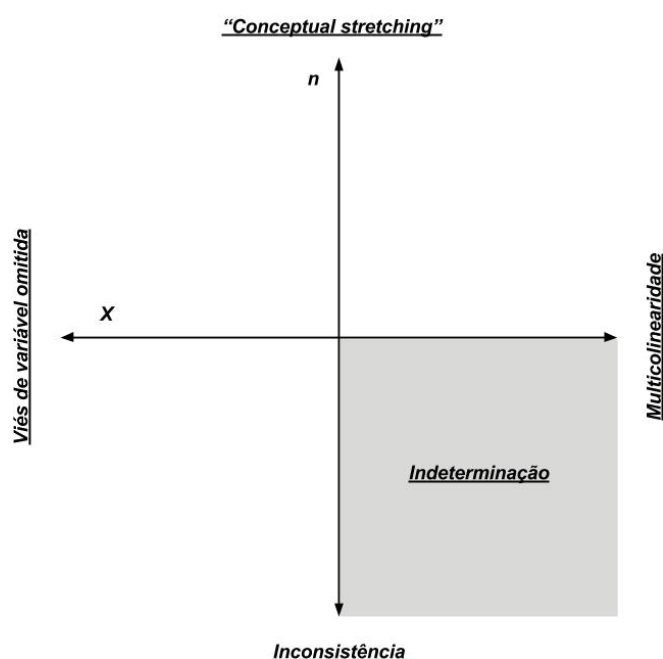
19 O número de graus de liberdade pode ser definido como o total de observações disponíveis menos o número de parâmetros que se deseja estimar a partir delas (WOOLDRIDGE, 2013, p. 100)

20 Importante notar que, dada a natureza dos dados utilizados em Ciências Sociais, sempre haverá alguma correlação entre as variáveis, de modo que a advertência é contra níveis elevados ou perfeitos de correlação (WOOLDRIDGE, 2013, p.84).

possível sem incorrer no viés de variáveis omitidas. Outra técnica é condensar algumas das variáveis do modelo em um indicador composto, através de técnicas de redução dimensional (análise fatorial, por exemplo) (FIGUEIREDO FILHO, SILVA, DOMINGOS, 2015).

Em síntese, o pesquisador frequentemente se depara com o quadro ilustrado pela Figura 4. O eixo horizontal representa o número de variáveis independentes do modelo e o vertical o número de casos. Estudos com poucas variáveis independentes (lado esquerdo) serão mais eficientes, porém com risco de viés, enquanto os à direita serão menos viesados, porém também menos eficientes, conforme o acréscimo de variáveis reduz graus de liberdade e aumenta as chances de multicolinearidade. Quanto ao N, vimos que KKV consideram mais consistentes (menor variabilidade) os estudos com muitas observações. Todavia, o aumento de observações, a depender do fenômeno, pode vir às custas da precisão conceitual (*concept stretching*). A indeterminação resulta da proporção entre variáveis mobilizadas e observações coletadas. Por isso, estudos com muitas variáveis e poucas observações (área cinza) padecem de indeterminação. Em síntese, tem-se um *trade off* entre precisão e viés (com respeito às variáveis) e entre consistência e exatidão conceitual.

Figura 4: Trade off na escolha entre número de observações e quantidade de variáveis



Fonte: elaborado pelo autor

3 CAUSALIDADE

A definição de causalidade empregada por KKV se assemelha ao modelo de Rubin (1974 *apud* KKV, 1994, p.82), no qual “o efeito causal é a diferença entre o componente sistemático das observações feitas quando a variável explicativa assume um valor e o componente sistemático das observações comparáveis quando a variável explicativa assume outro valor”. Tal definição indica que os autores aderem a uma compreensão contrafactual da causação – uma das quatro que exploraremos neste artigo.

A estimação de efeitos causais está assentada em algumas pressuposições que permitem conduzir tal raciocínio a despeito do “problema fundamental da inferência”, qual seja, que é impossível ver um fenômeno ocorrendo e não ocorrendo simultaneamente em uma unidade (REZENDE, 2017). A primeira suposição é a homogeneidade das unidades. Duas unidades são homogêneas quando seus valores esperados nas variáveis dependentes são os mesmos quando a variável independente assume certo valor. Como nem sempre é possível estabelecer essa equivalência no mundo real, supor um efeito causal constante para as diferentes unidades já poderia ser satisfatório. A segunda suposição é a independência condicional, isto é, os valores assumidos por Y não afetam os valores de X. A violação desse pressuposto é comumente apontada como uma das causas de endogeneidade.

3.1 Uma ou várias lógicas causais?

Em Ciência Política e Relações Internacionais, uma “causa” não significa sempre a mesma coisa para todos. Há apenas uma lógica pela qual se possa conceber que X cause Y? Ou ainda, o que significa dizer que X é a causa Y? A depender do autor, diferentes tipologias são propostas para classificar as lógicas causais²¹. Para os propósitos deste trabalho, destacamos as seguintes possibilidades:

1) Covariação/Correlacional: há causalidade quando vemos X e Y variarem conjuntamente. É o tipo de explicação que valida, por exemplo, métodos de variação concomitante, como o método da diferença de J. S. Mill (chamado de Método dos Sistemas Mais Semelhantes por Przeworski e Teune, 1970). Se existe uma associação forte entre variáveis teoricamente orientadas, espera-se que haja relação de causalidade entre elas — embora correlação não seja sinônima à causação, podendo antes ser uma correlação espúria.

2) Mecanismos: não se pode falar em causação apenas mensurando valores de duas variáveis ignorando o que acontece entre elas. Não há “ação à distância” (DAHL, 1957, p.204). É necessário abordar os mecanismos que conectam logicamente uma causa à sua consequência. Métodos como o *process tracing* e a análise histórica preocupam-se em desvelar este tipo de concatenação.

3) Experimental: causação está ligada à manipulação, ou seja, se o pesquisador pode manipular as variáveis, controlar pelos vieses, e observar efeitos, então poderá afirmar com segurança se há causação.

4) Contrafactual: responder à pergunta “como este caso se comportaria se este fator específico fosse removido?” é uma das formas de se poder asseverar uma relação causal. Todavia, esse tipo de raciocínio incorre problema fundamental da inferência. Por isso a importância da pesquisa comparada, já que ela se propõe a ser uma forma de controle em que se tenta mensurar os efeitos das variáveis em contextos diferentes (SARTORI, 1991).

Para alguns autores preocupados com os efeitos desse pluralismo causal sobre o *status* científico da Ciência Política, é preciso trazer esse sincretismo a uma fé única. Gerring (2005) afirma que embora a visão pluralista da causalidade tenha contribuído para dar complexidade ao “positivismo ingênuo”, também é problemática por exacerbar a diferenciação entre perspectivas que, na verdade, são complementares. Na sua visão, essa segregação compromete o objetivo de acúmulo de conhecimento científico. Ele propõe uma definição mínima, segundo a qual uma causa é um (ou vários) “eventos ou condições que aumentam a probabilidade de algum resultado ocorrer (sob condições *ceteris paribus*). X pode ser considerado uma causa de Y se (e somente se) aumenta a probabilidade de Y”²² (GERRING, 2005, p.169)

21 Ver Gerring (2005) para uma lista.

22 Tradução própria do autor a partir do original: “causes may be said to refer to events or conditions that raise the probability of some outcome occurring (under *ceteris paribus* conditions). X may be considered a cause of Y if (and only if) it raises the probability of Y.” (GERRING, 2005, p.169)

Entretanto, é curioso constatar que mesmo os projetos monistas terminam por fundar doutrinas separadas. Mahoney (2008), a seu turno, vê problemas em buscar a unificação recorrendo a noções probabilísticas. Isso pois encontramos na literatura metodológica distinção entre pesquisas orientadas a casos das orientadas a populações (RAGIN, 2004; COLLIER, BRADY, SEAWRIGHT, 2004; MAHONEY, 2008). Enquanto as primeiras querem explicar o que produz resultados específicos em contextos particulares, as últimas querem identificar efeitos médios ou marginais em populações. Isso tem consequências para a concepção de causação, pois, de um lado, se busca explicar resultados específicos e, de outro, probabilidades estatísticas. Mahoney (2008) entende que para conceber a causação de forma única deve-se conceder que a causa em nível individual é logicamente anterior à causa em nível populacional. Por isso, não se pode tomar como definição mínima “causa = aumento de probabilidade”, como é a proposta de Gerring (2005), já que não faz sentido falar de probabilidade quando $N = 1$. Portanto, Mahoney (2008) apresenta uma definição de causalidade em nível individual em termos de condições necessárias e/ou suficientes.

Além das combinações tradicionais de análises que lidam com causas necessárias e/ou suficientes (1. Causa necessária e suficiente; 2. Causa necessária mas insuficiente; 3. Causa suficiente mas desnecessária) ele propõe mais duas (derivadas de MACKIE 1965 *apud* MAHONEY, 2008): (4) Causas INUS (*Insufficient but Necessary part of an Unnecessary but Sufficient condition*), para as quais não existe uma única causa necessária ou suficiente. A causa individual é na verdade uma parte insuficiente, mas necessária, de um arranjo que, por sua vez, é desnecessário, mas suficiente.

Por exemplo, tome-se:

$$Y = (A \ \& \ B) \mid (C \ \& \ D).$$

Em termos formais, “A” é insuficiente por precisar estar associada com “B” (“&” representa o “e” lógico), mas este par não é necessário, porque se ocorrer o par “C & D”, o fenômeno “Y” ocorrerá assim mesmo (“|” representa o “ou” lógico). Já as (5) Causas SUIN (*Sufficient but Unnecessary part of a Insufficient but Necessary context*) são aquelas em que a causa individual é parte suficiente, mas desnecessária, de um fator, e este fator é por sua vez causa insuficiente e necessária de um fenômeno. Em termos formais, suponhamos:

$$Y = X \ \& \ Z; \text{ onde: } Z = A \mid B.$$

“A” é suficiente, mas não necessário, para produzir “Z”, e este é insuficiente, mas necessário, para produzir “Y”. Logo, “A” (e “B” também) é para “Y” uma causa SUIN.

Para unificar essa tipologia determinística com a visão probabilística de causação, o autor postula que causas ocorrem em nível populacional seguindo uma lógica INUS, já que a heterogeneidade crescente do nível populacional de análise torna mais raras as causas (5) SUIN, (1) necessárias e suficiente, e (2) necessárias e insuficientes.

Uma das consequências das causas INUS é a equifinalidade, ou seja, diferentes grupos de variáveis levam ao mesmo resultado (os pares “A & B” ou “C & D” no primeiro exemplo). Essa característica está tacitamente prevista em modelos lineares, porque esses geralmente têm muitas variáveis, e há virtualmente infinitas combinações de valores delas que darão um mesmo escore numérico em Y. Isso significa que nenhuma causa no modelo é necessária, mas diferentes valores de variáveis (associados com o termo de erro) são suficientes para produzirem o resultado²³.

23 Em aprimoramento posterior deste modelo, Mahoney, Kimball e Koivu (2009) buscam solucionar um desafio das pesquisas não estatísticas: se múltiplas causas estão interagindo na produção de um fenômeno, como estabelecer qual a mais importante? Enquanto regressões lineares podem responder essa pergunta

3.2 Pressupostos ontológicos: Probabilidade x Determinismo

Uma das principais divisões entre tradições de pesquisa em Ciência Política e Relações Internacionais se ancora em como o pesquisador entende que a realidade se comporta²⁴. Ou o mundo é probabilístico e, portanto, ainda que se coletem todas as variáveis independentes possíveis para explicar um fenômeno, não se pode dizer ao certo o que ocorrerá, pois o acaso sempre desempenha algum papel; ou o mundo é determinístico: as coisas acontecem ou não, *tout court*, e destarte a aparente aleatoriedade é fruto de alguma variável independente que ainda não se coletou.

KKV (1994) adotam uma visão probabilística. Eles recomendam, portanto, diferenciar entre componentes sistêmicos (i.e.: que devem ter efeito causal regularmente explicado pela teoria) e não sistêmicos (i.e.: efeitos não contemplados pela teoria; ruído) com base em critérios *a priori* sobre previsibilidade e aleatoriedade. Por isso que eles também, embora reconheçam que o critério popperiano da falseabilidade seja central para a empreitada científica, o relativizam. Eles entendem que a falsificação de certa teoria por novas evidências é mais um indício dos limites da aplicabilidade da mesma do que uma sentença condenando-a ao descarte²⁵.

Já Ragin (2004, p.124) entende que pesquisas orientadas a casos e pesquisas orientadas a variáveis são “modos diferentes de pesquisa com suposições operacionais distintas”. Em específico, há preocupações práticas que as separam: (1) a constituição de casos: estes são entendidos, na pesquisa orientada a casos, não como observações homogêneas de um *pool*, mas configurações complexas de eventos. (2) Definição de casos negativos: estudos de caso são frequentemente acusados de escolherem apenas segundo valores positivos de suas variáveis dependentes, o que traz problemas para a generalização. Segundo Ragin (2004), a preocupação com a generalidade já é contemplada durante o próprio processo de seleção de casos, quando se mobilizam implicitamente preocupações sobre condições necessárias e suficientes para a produção do fato. (3) Análise de causas conjunturais e múltiplas: pesquisas orientadas a casos veem causação como dependendo de conjunturas, portanto não se prestam a medir o efeito independente de um X sobre Y, como podem as pesquisas orientadas a variáveis. Por fim, (4) o tratamento de casos desviantes: enquanto a pesquisas orientadas a variáveis tendem a considerá-los como resíduo ou ruído, as orientadas a casos buscarão explicá-lo em termos substantivos, o que é o mesmo que dizer que as duas abordagens têm orientações distintas sobre o que se deve fazer com o “erro” de um modelo teórico.

Essas concepções diferentes afetam, por exemplo, a validade de certos tipos de desenho de pesquisa. Para uma visão determinística, que concebe um universo causal invariante, um único estudo de caso pode ser o suficiente para desbancar uma teoria (ROGOWSKI, 2004). Probabilisticamente, contudo, os achados de um estudo de caso devem ser interpretados com mais parcimônia.

4 CONCLUSÃO

O presente trabalho buscou apresentar, de forma introdutória e didática, as principais questões acerca de desenhos de pesquisa em Ciência Política e Relações Internacionais. Como foi evidenciado, a nossa disciplina coleciona um arcabouço robusto de recomendações sobre o que fazer para melhorar desenhos de pesquisa, inferências e reivindicações

observando, por exemplo, o valor dos coeficientes ajustados, sabendo assim qual variável produz maior efeito sobre Y (vide a respeito FIGUEIREDO FILHO et al, 2011, p.49), para análises históricas eles propõem o que chamam de “método de elaboração de sequência”, em que a *identidade* entre variável independente e dependente pode ser usada como critério de relevância.

24 Na divisão apresentada por Almond (1988) entre ideologia (esquerda/direita) e método (*hard/soft*), entra também a problemática de qual a posição do pesquisador sobre a possibilidade de produção de conhecimento científico objetivo. Esta dimensão política da disciplina não é abordada no presente artigo.

25 Um exemplo seria a teoria da escolha racional, que, apesar das crescentes evidências contrárias aos seus postulados, não chega a ser julgada obsoleta ou descartável, pois ainda permite explicar boa parte do comportamento humano.

de causalidade. Viu-se que alguns elementos são terreno comum aos muitos praticantes, como a importância de se escolherem perguntas de pesquisa relevantes e o zelo com a validade dos conceitos e confiabilidade dos dados. Outros pontos, em contraste, são menos consensuais.

Em particular, foi possível constatar que as divergências indicam duas tradições polarizadas. De um lado, pesquisas *large n*, quantitativas, voltadas à população/variáveis, preocupadas com mensurar os efeitos das causas e impactos marginais, com uma visão de mundo probabilística, e, de outro, estudos *small n*, qualitativos, voltados a casos, preocupados com descobrir as causas dos efeitos, e ontologicamente deterministas. Poder-se-ia acrescentar que os primeiros estão mais preocupados em alcançar a generalidade enquanto os segundos com a explicação de fenômenos particulares, mas um dos pontos de tensão é precisamente o quanto esta última tradição também é válida para a produção de hipóteses universalizáveis. Enquanto alguns autores enxergam complementaridade entre os campos, outros veem excepcionalidades que apontam para lógicas operacionais distintas.

Em conclusão, parece-nos importante destacar a ênfase dada por KKV na ancoragem das generalizações em observações: isso separa ciência de especulação. Todavia, o cientista *político* tem diante de si um objeto peculiar. Nem todos os fenômenos de interesse para Ciência Política e Relações Internacionais são “geradores de observações” de alta frequência. Alguns produzem dados em profusão e com regularidade (e.g.: votações, fluxos financeiros, tratados comerciais, censo, legislação, etc), ao passo que outros só ao longe, raramente e de relance se deixam ver (legitimidade, ideologia, barganhas, coerção). Em se tratando de política, a depender da área, a parcela de informação que vem à luz é inferior à que permanece na sombra. Assim, é salutar, por um lado, manter ambições humildes sobre o real alcance das nossas observações em se tratando de processos políticos; e, por outro, se precaver contra a cilada de considerar alguns estudos mais sofisticados pelo fato de terem tão somente elegido um fenômeno que, por sua constituição, gere uma riqueza maior de dados.

O que resulta claro dessa exposição é que a atividade científica envolve *trade offs* e que cabe ao pesquisador conhecer com propriedade suas próprias indagações, o assunto que deseja investigar, as possibilidades e limitações que a realidade empírica impõe a esse projeto, e assim optar pelo ferramental apropriado para cumprir este desígnio. A fronteira inferencial será sempre determinada pela realidade fática e o quanto ela nos dá para trabalhar.

Com este artigo, esperamos apresentar aos acadêmicos em formação as principais balizas do debate em desenho de pesquisa em Ciência Política e Relações Internacionais atualmente. Tal conhecimento é um ponto de partida fundamental tanto para a capacidade de interação crítica com a literatura produzida por seus pares, quanto para escolha consciente de como melhor tratar um problema de pesquisa.

REFERÊNCIAS

- ALMOND, Gabriel A. (1988). "Separate tables: schools and sects in political science" *PS: Political Science and Politics*. Vol. 21, No. 4: 828-842
- ALMOND, Gabriel A.; GENCO, Stephen J. (1977) "Clouds, clocks and the study of politics". *World Politics*. Vol. 29, No. 4: 489-522
- BARBOZA, Danilo P.; GODOY, Samuel R. (2014) "Superando o 'calcanhar metodológico'? Mapeamento e evolução recente da formação em métodos de pesquisa na pós-graduação em Ciência Política no Brasil". *Trabalho apresentado no IV Seminário Discente da Pós-Graduação em Ciência Política da USP*. 07 a 11 de abril de 2014.
- BLALOCK Jr, H. M. (1961). "Theory, Measurement, and Replication in the Social Sciences". *American Journal of Sociology*, Vol. 66, No. 4: 342-347
- BRADY, Henry; COLLIER, David. (2004). *Rethinking social inquiry: diverse tools, shared standards*. Maryland: Rowman & Littlefield Publishers.
- COLLIER, David; BRADY, Henry; SEAWRIGHT, Jason. (2004), "Sources of leverage in causal inference: towards an alternative view of methodology" in Henry Brady; David Collier (eds). *Rethinking social inquiry: diverse tools, shared standards*. Maryland: Rowman & Littlefield Publishers, pp. 229-266
- COLLIER, David; MAHON, James E. (1993). "Conceptual "stretching" revisited: adapting categories in comparative analysis." *American Political Science Review*, Vol. 87, No. 4: 845-855
- DAHL, Robert A. (1957). "The Concept of Power". *Behavioral Science*. Vol. 2, No. 3: 201-215
- FIGUEIREDO FILHO, Dalso; SILVA, Lucas; DOMINGOS, Amanda. (2015). "O que é e como superar a multicolinearidade? Um guia para a ciência política." *Conexão Política*. Vol. 4, No. 2: 95-104
- FIGUEIREDO FILHO, Dalson, et al. (2011), "O Que Fazer e o Que Não Fazer Com a Regressão: pressupostos e aplicações do modelo linear de Mínimos Quadrados Ordinários (MQO)." *Revista Política Hoje*, Vol. 20, No. 1
- FREITAS, Vitor S.; NETO, Fernando B. (2016). "Qualitative Comparative Analysis (QCA): usos e aplicações do método". *Revista Política Hoje* Vol. 24, No. 2: 103-118
- GEDDES, Barbara. (1990), "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics." *Political Analysis*. Vol. 2, No. 1:131-150
- GERRING, John. (2004), "What is a case study and what it is good for?" *American Political Science Review*, Vol. 98, No. 2: 341-354.
- GERRING, John. (2005), "Causation: a unified framework for the social sciences." *Journal of Theoretical Politics*, Vol. 17, No. 2: 163-198

- KENNEDY, Peter. (2003). *A guide to econometrics*. Cambridge: MIT Press.
- KEOHANE, Robert O.; NYE, Joseph N. (2001). *Power and interdependence*. New York: Pearson.
- KING, Gary. (1995), "Replication, replication." *PS: Political Science & Politics*. Vol. 28, No. 3: 443–499.
- KING, Gary. (2015) [1995]. "Replicação, replicação". *Revista Eletrônica de Ciência Política*, Vol. 6, No. 2: 382-401
- KING, Gary; KEOHANE, Robert O.; VERBA, Sidney. (1994), *Designing Social Inquiry: scientific inference in qualitative research*. Princeton: Princeton University Press.
- LIJPHART, Arend. (1971), "Comparative politics and the comparative method." *American Political Science Review*, Vol. 65, No. 3: 682-93
- MAHONEY, James. (2008), "Towards a unified theory of causality." *Comparative Political Studies*. Vol. 41, No. 4-5: 412-436
- MAHONEY, James; KIMBALL, Erin; KOIVU, Kendra L. (2009), "The Logic of Historical Explanation in the Social Sciences." *Comparative Political Studies*, Vol. 42, No. 1: 114-146
- MEDEIROS, Marcelo de Almeida et al. (2016). "What does the field of International Relations look like in South America?" *Rev. bras. polít. int.*, Vol. 59, No. 1, e004, 2016
- MERTON, Robert K. (1942). *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press.
- MILLER, Bernhard. (2011), "Making measures capture concepts: tools for securing correspondence between theoretical ideas and observations." in Thomas Gschwend; Frank Schimmelfennig (orgs.). *Research Design in Political Science: how to practice what they preach*. London: Palgrave Macmillan, pp.83-102
- MUNCK, Gerardo. (2004), "Tools for qualitative research." in Henry Brady; David Collier. *Rethinking social inquiry: diverse tools, shared standards*. Maryland: Rowman & Littlefield Publishers, pp. 105-122
- NEIVA, Pedro. (2015). "Revisitando o calcanhar de Aquiles metodológico das ciências sociais no Brasil". *Sociologia, Problemas e Práticas*. No. 49: 65-83
- PEARSON, Karl. (1900) *The Grammar of Science*. Londres: Adam and Black. Disponível em: < <https://archive.org/details/grammarofscience00pearuoft> > Data de acesso: 12/08/2017
- PÉREZ-LIÑÁN, Aníbal. (2010), "El método comparativo y el análisis de configuraciones causales." *Revista Latinoamericana de Política Comparada*. Vol. No. 3: 125-148
- POPPER, Karl. (2002) [1935]. *The logic of scientific discovery*. Londres: Routledge.

- PRZEWORSKI, Adam; TEUNE, Harry. (1970), *The Logic of Comparative Social Inquiry*. New York: Wiley-Interscience.
- RAGIN, Charles. (2004), "Turning the tables: how case-oriented research challenges variable-oriented research." in Henry Brady; David Collier (eds.). *Rethinking social inquiry: diverse tools, shared standards*. Maryland: Rowman & Littlefield Publishers, 123-138
- REZENDE, Flávio da Cunha. (2017). "O Pluralismo Inferencial na Ciência Política Pós-KKV (2005-2015): Argumento e Evidências". *Revista Política Hoje*. Vol. 26, No. 1: 241-278
- ROGOWSKI, Ronald. (2004), "How inference in the social (but not the physical sciences) neglects theoretical anomaly". In Henry Brady; David Collier. *Rethinking social inquiry: diverse tools, shared standards*. Maryland: Rowman & Littlefield Publishers, pp. 75-84
- SARTORI, Giovanni. (1991), "Comparing and miscomparing." *Journal of Theoretical Politics*. Vol. 3, No. 3: 243-257
- SILVA, Glauco Peres da. (2015). "Desafios ontológicos e epistemológicos para os métodos mistos na ciência política". *Rev. bras. Ci. Soc.* Vol.30, No.88: 115-128
- SOARES, Gláucio. (2005), "O calcanhar metodológico da ciência política no Brasil." *Sociologia, problemas e práticas*, No. 48: 27-52
- TARROW, Sidney. (2005), "Bridging the quantitative-qualitative divide." in Henry Brady; David Collier. *Rethinking social inquiry: diverse tools, shared standards*. Maryland: Rowman & Littlefield Publishers, pp.171-180
- WOOLDRIDGE, Jeffrey M. (2013). *Introductory Econometrics: a modern approach*. 5th edition. Ohio: South-Western.

Resumo

O artigo apresenta uma introdução didática sobre desenhos de pesquisa, inferência e causalidade em Ciência Política e Relações Internacionais. Destina-se, portanto, a familiarizar estudantes do campo com o ferramental desta atividade científica, bem como introduzi-los aos principais debates que perpassam a disciplina. Para tanto, revisamos os elementos centrais do livro *Designing Social Inquiry* e as publicações que responderam a ele, endossando ou contestando as suas concepções sobre as melhores práticas científicas. Com este artigo esperamos equipar os alunos de Ciência Política e Relações Internacionais para o pensamento crítico sobre as escolhas envolvidas na elaboração e execução de um desenho de pesquisa.

Palavras chave: ciência política, metodologia, inferência, causalidade, KKV

Abstract

This article seeks to present, in a pedagogical and introductory way, central issues concerning research design, inference and causality in Political Science and International Relations. It aims, therefore, at familiarizing students of this field with the common tools of this scientific endeavor, as well as introducing them to the main debates that cut across the discipline. In order to do so, we review the guidelines presented in the book *Designing Social Inquiry* and in the publications that answered it, either endorsing or contesting its conceptions on the best scientific practices. The objective of this contribution is to better equip Political Science and International Relations students to think critically about the choices involved in the elaboration and execution of a research design.

Key words: political science, methodology, inference, causality, KKV

Research design, inference and causality in Political Science and International Relations: a didactic introduction

Rafael Mesquita
Departamento de Ciência Política - UFPE

introduction

One of main issues political scientists and internationalists face today in Brazil is the lack of appropriate understanding of research design and its accompanying features: the appropriate use of research methods and understanding of inference and causality. Soares (2005) warned us about the striking negligence and unpreparedness in Brazilian Social Sciences regarding these topics. Ever since this first diagnosis, we have seen reasonable progress concerning the proliferation of methodology courses at universities (BARBOZA & GODOY, 2014)¹. However, even after a decade, publications with methodological rigor in Political Science and International Relations are few (NEIVA, 2015; MEDEIROS et al. 2016).² Therefore, it is still an important task to train those who have just joined this field of study in how to adequately handle its tools.

With this in mind, the current paper aims to provide a didactic introduction to the main discussions in the literature about methodology in Political Science and International Relations. We will dedicate special attention to reviewing the guidelines established by canonical authors in the field³, as well as the improvements and counter-arguments developed by other researchers about such guidelines. It is, therefore, inspired by the discussion started by King, Keohane and Verba (1994) in *Designing Social Inquiry* and its repercussions. It's been over twenty years since the first edition of the book came out and left a strong mark in the field, shaping the research agenda in regards to research design and inference (REZENDE, 2017).

In this article, we seek to explain the concepts of (1) *research design*, (2) *inference* and (3) *causality*, being the text also structured along these three topics.

1 Barboza and Godoy's (2014) study shows that, between 1998 and 2012, universities such as UFMG, UFPE and IUPERJ expanded their offer of courses on methodology. Neiva (2015) also remarks that authors connected to these institutions, alongside with UFPR, have published more articles using quantitative methods between 1997 and 2012.

2 Neiva (2015) points out that the use of advanced statistics in Social Sciences articles in Brazil has been growing (going from 3.3% in 1997 to 7.6% in 2012), but the field remains mainly "non-quantitative". Medeiros et al. (2016) show that less than 13% of International Relations articles in South America even mention some sort of methodology explicitly.

3 It is important to point out that the main authors cited here have a positivist approach to science, which means that this analysis will exclude post-positivist authors, who questioned the possibility of an objective approach to scientific knowledge, dissociated from historical processes, power relations and ideologies (ALMOND, 1988).

1 RESEARCH DESIGN

Rezende (2017, page 248) defines research design as “identification strategies, where premises that allow us to identify specific causal effects are established, and which contain formal proofs that make it possible for these to lead to causal effects in infinite samples”. This conceptualization offers us a starting point, pointing out the most important aspects of a research design: identification, causality and formal elaboration.⁴

In order to understand how research design leads to scientific knowledge production, this section will cover (1) the criteria which grant the scientific status to Political Science and International Relations research; (2) the elements that must be present in research design; (3) around which aspects different research designs tend to disagree and establish different traditions; and (4) some common issues around research design.

1.1 What makes a research scientific?

What makes scientific inquiry different from other forms of knowledge? This question has a long tradition in the philosophy of science. Over a century ago, Karl Pearson (1900) wrote that the unity of all science lies in its methods, not its material. In 1942, Robert Merton stated that scientific knowledge must be defined by its communal character, for being universal, impersonal and for always being subjected to “organized skepticism”.

In *Designing Social Inquiry*, King, Keohane and Verba (1994) – henceforth KKV – define scientific research according to four criteria: (1) its objective is to reach causal inference or descriptive inference; (2) its procedure is public and it is transparency that will allow the whole scientific community to judge its validity, learn from its method, replicate its result, understand its limitations in order to overcome them and make comparisons with different researchers⁵; (3) its conclusions are uncertain, given that inference from imperfect data can only lead to less than perfect conclusions; and (4) its content is the method, that is, the scientific status of a research lies in the methods and rules utilized, not in its subject matter, which is why we can “science” call such a multifaceted practice, encompassing many different areas.

1.2 What must a research design have?

KKV (1994) define research design as the combination of four elements. The first element is the *research question*, which must be relevant for both society and academia. That means that its findings must actually have an impact on people’s lives. Moreover, the question must contribute to the scientific state of the art. Therefore, it needs to engage with the existing literature and seek to advance existing knowledge. The researcher must therefore be cautious so as not to explore a relevant topic (e.g.: hunger, war, poverty) without the systematic approach that makes the investigation properly scientific, also while avoiding scholarly puzzles of little relevance outside of the walls of academia.

The second element is *theory*, defined by KKV as a rational and precise speculation about the possible answers for the research question, and that also includes a declaration on why that specific answer is the correct one. *Falsifiable* theories are properly scientific. This recommendation comes from Karl Popper, who in 1935 argued that knowledge systems must be falsifiable by experience, in order to be considered empiric and scientific. Therefore, the empiric method fulfils its role when it exposes a system to the possibility of being contradicted. Because of that, it is recommended to select theories that yield the maximum possible amount of observable implications. Thus, it is possible to

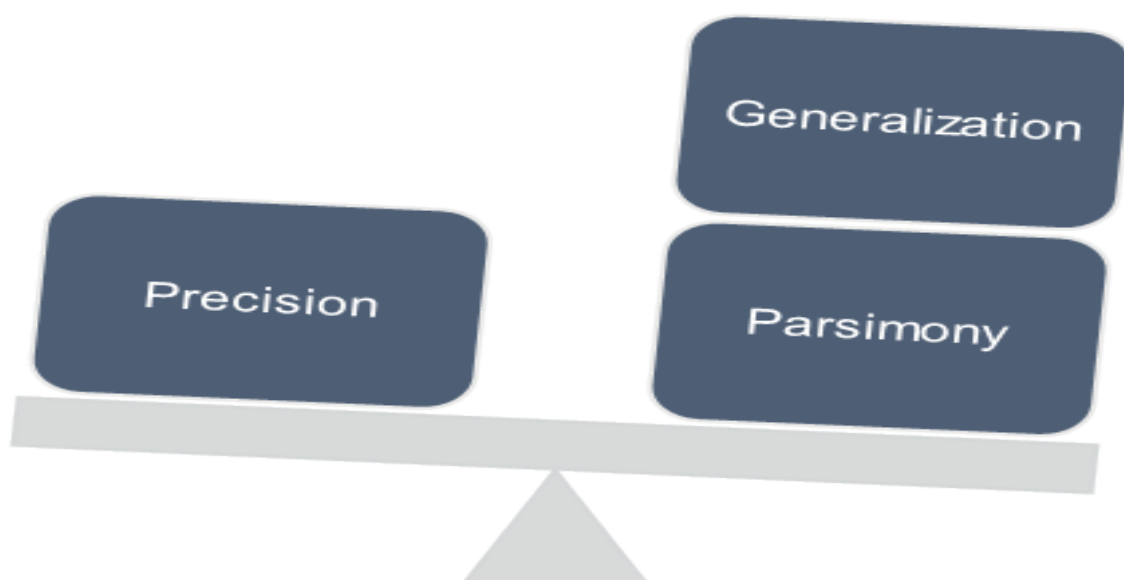
⁴ Nevertheless, it should be pointed out that research design shouldn’t be just about causality: research geared towards descriptive inference should also mobilize it in their identification efforts.

⁵ King (2015) [1995] stressed later on the importance of making replication a standard practice for Political Science. The *Revista Eletrônica de Ciência Política* translated the original version of the paper, published on *PS: Political Science and Politics*, which had a special edition about Research Methodology. Also in 2015, *Revista Política Hoje* edited a special edition about the same theme (<https://periodicos.ufpe.br/revistas/politica hoje/issue/view/722/showToc>).

optimize the opportunities to verify the empirical implications of these theoretical conjectures and thus to confirm or refute them.

Przeworski and Teune (1970) also seek to define standards to assess theories in comparative research, suggesting four cardinal objectives: (1) precision (the level of detail reached in the explanation of the cases); (2) generalization (how broad the explanation is); (3) parsimony (being able to explain a high number of cases using only a small group of causes) and (4) causality (establishing a relationship of cause and effect between the variables used). It is not possible to follow all objectives at the same time and with the same success. Researches that try to be parsimonious and general are usually not as precise, for using a small number of variables allows to deal with a high number of cases, however, as cases increase and explanatory factors decrease, the smaller the amount of detail afforded. Alternatively, more precision demands greater detailing of specific cases, meaning that studies which are more precise will be less general and less parsimonious. There is a tradeoff between those goals and it is up to each scientist to choose which one is more important to their endeavor. Figure 1 illustrates the relationship between precision and generalization.

Figure 1: Tradeoff between Precision vs. Generalization and Parsimony



Source: elaborated by the author, based on Przeworski and Teune (1970)

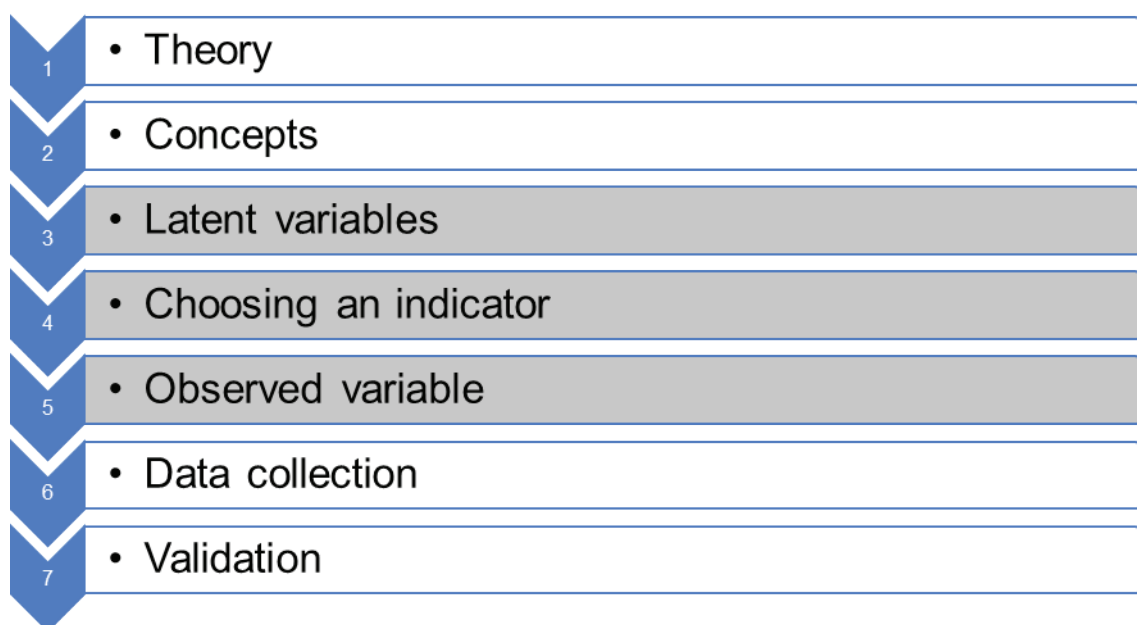
The third component KKV point out is the empirically collected *data*, which will test the hypotheses proposed by the theory. The unavoidable step between theory and data is called *measurement*. For Miller (2011) this is the heart of research design, for it allows the theoretical and the empirical to come together.

What is, then, “measuring”? Miller proposes an analogy with the natural world: suppose we have an abstract concept we would like to measure (temperature, for example), however, it is impossible to observe it directly. Knowing that the volume of mercury is affected by changes in temperature (observable implication), we can use it to measure temperature. The option to use mercury is called *operationalization*. Once measurement has been taken, it is time to verify if its result really does correspond to the change in temperature, which brings us to the issue of *validation*, that is, the registered measure matches the idealized concept. Finally, it is important to make sure that other researchers will achieve the same results, under the same conditions. This is called *reliability*.

Moving from nature into politics, we see the same trend, going from abstract concepts to accessible implications. Dahl (1957), when referring to “power” – a concept that is as omnipresent in Political Science as it is cloudy – he starts off by defining it as follows: A has power over B if it can make B do something B wouldn’t otherwise do. Here we have a concept that produces an observable implication (the change in B’s behavior) and that can therefore be operationalized (e.g.: seeing how a congressman B votes after interacting with A, his party leader). In the field of International Relations, Keohane and Nye (2001) state that interstate relations are now marked by interdependence, which can be observed in two elements: sensitivity (being exposed to costly effects due to changes originating abroad) and vulnerability (the cost to adapt to the changes originated abroad). The sensitivity and vulnerability of a country, for example, in the field of energy resources, can be operationalized by looking into how much oil a country imports and whether it is possible to implement policies that would raise or lower this amount.

Putting it into steps, social scientists should: (1) start from a theory in order to (2) develop clear and precise concepts about what they want to measure. From these concepts, we will derive (3) the latent variables, which are those that the researcher would ideally like to observe⁶. As we frequently study phenomena which cannot be directly observed, we resort to (4) an indicator that points to (5) the observed variable. The operationalization process encompasses steps 3 to 5. After (6) data gathering, the researcher should ensure that (7) they are valid and reliable. This whole procedure is summed up on Figure 2.

Figure 2: Steps of the measuring process



Source: elaborated by the author, based on Miller (2011). Grey cells represent the operationalization stage.

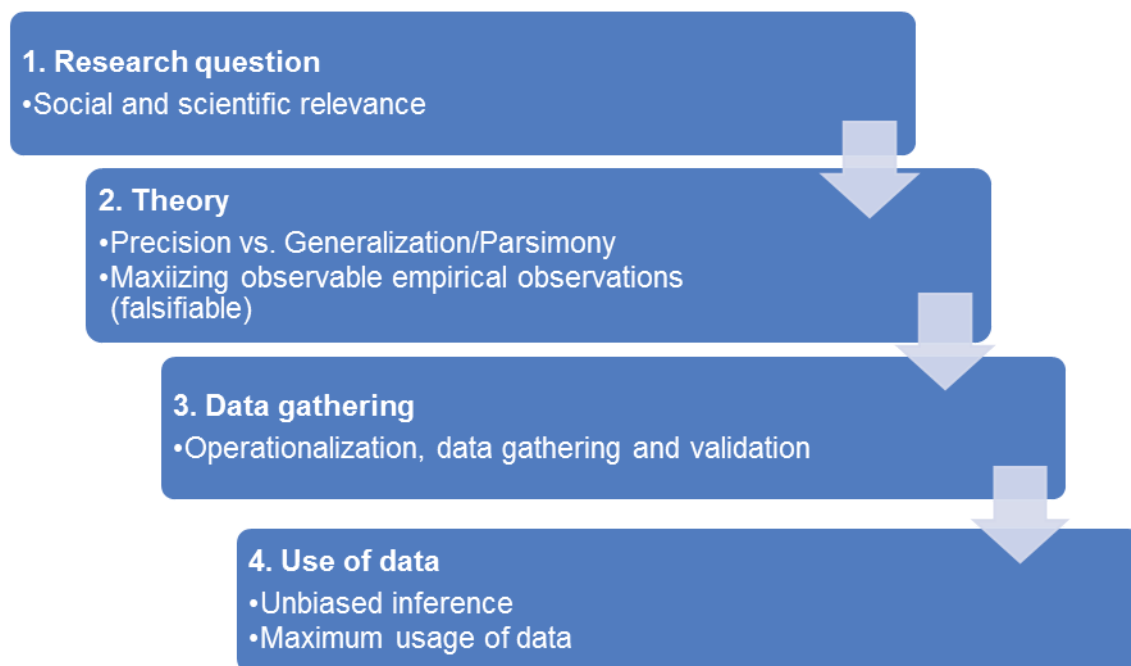
However, it is important to keep in mind that validity and reliability can be quite fleeting outside a purely experimental context – which is often the case for Social Sciences. This is because the researcher, hoping to find in the empirical world some concept (e.g.: “racial discrimination”), collects the values of some proxy (“income”), but these values are generated not only by the concept of interest, but rather by a myriad of confounders (“level of education”, “ability”, etc). Because of

⁶ Miller (2011) states that latent variables are not synonymous to concepts, because some concepts require more than one latent variable to be measured.

that, Blalock (1961, page 347) states that “a satisfactory measure of a theoretically defined concept calls for a theory of how other variables are operating” (BLALOCK, 1961, page 347).

Finally, the fourth and last element of research design defined by KKV is the *use of data*. Data is best used when it is utilized to make unbiased inferences and the harnessing of information acquired is maximized – issues which will be better detailed on the section about inference and causality. To sum up, the steps defined by the authors can be synthesized as follows.

Figure 3: Elements of a research design



Source: Elaborated by the author, based on Przerworski and Teune (1970), KKV (1994) and Miller (2011)

1.3 Differences between research designs

There are many questions in Political Science and International Relations, as well as many different ways of answering them. Because of that, the research designs available for scientists are also plentiful, each with its advantages and limitations. There are many ways to differentiate between research designs, but this paper will focus on the following elements: (a) methodology, (b) number of observations and (c) the use of experiments or lack thereof, given that these matters engender important consequences in terms of justifying the choice of one research design over the other, inferential validity and causality.

a) Methodology: Qualitative x Quantitative

The division between quantitative and qualitative methods has been a theme of discussion for many years in Social Sciences and was one of the reasons behind the development *Designing Social Inquiry*⁷. The work stated that, according to the authors, both methods differ only in style but share the same inferential logic; and the book aimed to

⁷ See also Almond (1988) on the methodological and ideological divide of US Political Science.

assist qualitative researchers (who, according to the authors, tend to be less methodologically rigorous) to be more aware of their research design and thus learn how to boost the validity of their findings (KKV, 1994). Much of the criticism directed at KKV's project were condensed on *Rethinking Social Inquiry* by Brady and Collier (2004), who stated that the list of good practices offered was, in fact, based on quantitative references, aimed at linear regressions and therefore neglected the virtues specific to qualitative methods, such as the potential these types of studies have for generalization and creating *post hoc* hypotheses. Some of the authors featured on Brady and Collier's book shared with KKV the understanding of the existence of only one inferential logic for both methods and worried about correcting and perfecting the recommendations to combine both traditions. Tarrow (2004), for example, lists a toolbox of strategies that might be used to combine quantitative and qualitative methods. Other authors, however, consider that the gap between the methods is larger and insisted that qualitative logic is specific and does not allow concepts to travel from one side to the other. This exceptionalist view was based firstly on the number of observations needed to propose a causal relation and secondly on ontological considerations about the nature of reality: whether it is deterministic or probabilistic.⁸

b) Number of observations: Large n x Small n

Large n research is (almost) exclusively quantitative, simply because statistic methods are more feasible to embrace a large number of observations. Because it deals with abundant cases, it naturally loses in richness of detail and depth, but it gains in generalizability and, most importantly, in inferential leverage concerning the validity of the studied relationship. Thus, explanations afforded by these studies do not have a great degree of specificity, meaning that they do not exhaustively explain the causes involved in producing a specific phenomenon, being better suited to explain the effects of a causal relation across a population. They are, thus, recommended to measure the *effects of the causes* (i.e.: "how much does X affect Y?" or "how much of Y can be explained by X?"), being therefore also called "variable oriented" or "population oriented" (MAHONEY, 2008. RAGIN, 2004).

The chief way of conducting large n studies is through ordinary least square (OLS) regression, which alone represents almost a third of all publications in Political Science in the United States (KRUEGER & LEWIS-BECK, 2008 as cited in FIGUEIREDO FILHO et al, 2011, page 46). Regression analysis consists of investigating a possible causal relation between variables. More specifically, it allows us to gauge the association between a dependent variable and one or more independent variables, its magnitude (weak or strong) and direction (positive or negative). The value of this kind of analysis lies, firstly, in its straightforward interpretation and, secondly, in its robustness against the violation of certain assumptions, and also because it allows the researcher to explain, interpret and predict values for the variables of interest (FIGUEIREDO FILHO et al. 2011, pages 48-49).⁹

As for small n researches, they deal with a smaller number of cases. They rely mainly on qualitative methods and gain in depth and contextual knowledge what they lose in parsimony and generalization. Because of the intense focus on few cases, it is the kind of research that is concerned with the production of specific outcomes. To put it simply, it is centered on the *causes of the effects* (i.e.: "what causes Y?"), which is why they are also referred to as case oriented (MAHONEY, 2008; RAGIN, 2004).

Regarding the number of observations, case studies deserve special attention. These are traditionally thought of as small n studies where n=1 (LIJPHART, 1971). Gerring (2004) suggested redefining them as the intense study of one specific unit with the objective of understanding a bigger group of similar units. Since he defines units as possibly

⁸ See also Silva (2015) for a discussion about the challenges of mixed methods, regarding (i) philosophical stance taken in regard to methodology and expectations of reality; and (ii) methodological pluralism within qualitative methods.

⁹ See also chapter two of Wooldridge (2013) for a practical introduction.

containing various other cases, the author understands that case studies might be large *n* as well as small *n* in nature. Its logic is to detect covariation, which can be found both within a unit, across units, combining both, statically or across time for all units, which leads to different types of study.

For Lijphart (1971), the case studies that add the most to theory are those which seek, through observations from one unit, to formulate propositions that might be applicable to more units (“hypotheses-generating studies”), and those that investigate deviant cases, which are distant from the pattern predicted by theory and, because of that, might contain important information about the causal relation at hand (“deviant case studies”). Gerring (2004) adds that, the smaller the number of empirical cases with useful variation on the relevant variable, the more relevant will case studies be.

c) Experiment x Observation

Experimental studies seek to find out if *X* causes *Y* by way of deliberate manipulation by the researcher: one group will be exposed to the possible cause while the other will not (“treatment” and “control” groups, respectively). The average difference between the values of the dependent variable between the groups will be called the cause. The main advantage of this kind of design is that it allows the researcher to determine the values of the independent variable and, specifically, to select the observations, which allows for the usage of randomization as a way to ensure that the selection procedure is not correlated to any other variable, which keeps a host of biases away.

As for observational studies, there is no way for the researcher to control and manipulate the context. He/she faces a universe that has been molded by history and/or nature. Therefore, the analyst needs to use a toolbox of techniques to establish causality, observing in the empirical world the many possible forms the relation between *X* and *Y* might take, seeking thus to make inferences about their interaction. While the use of experiments is the rule for the Natural Sciences, they are rare in Political Science and International Relations for practical as well as ethical reasons.¹⁰

1.4 Some research design issues: selection bias and omitted variable bias

A recurring issue, whose treatment is still a controversy in the scientific community, is selection bias. This happens when the selection of cases is affected by the dependent variable, that is, the only cases studied are those where the phenomenon of interest occurs. As shown by Geddes (1990), this can lead to an underestimation of the causal effect. Even if it seems intuitive to only study positive cases (e.g.: if I want to better understand electoral victories, why would I include the losing candidates in the study?), this actually reduces the variance of the dependent variable (COLLIER & MAHONEY, 1996). Selection that is based on the independent variable, alternatively, does not create a bias on results. Although selection bias is usually the result of conscious or subconscious choices made by the researcher, some biases are a result of historical or natural action, which privileged the appearance of certain phenomena while suppressing others.

To some authors from the case study tradition, however, these warnings must be reviewed. Brady and Collier (2004) state that the warnings made by KKV against selection biases, although useful, are only applicable to qualitative cross-case research. As for in-depth studies of a single unit, also known as within-case studies, the truncation of the

¹⁰ Nonetheless, the use of “semi-experimental” designs has become more popular. According to Google Scholar data, only 338 articles with the keywords “quasi-experiment” and “political science” could be found in 2005. In 2015, there were 1,520 (research done by the author on scholar.google.com.br). These are also called “quasi-experimental” or “natural experiments” depending on the case, and typically they do not allow the researcher to guarantee randomized samples or to have control over the application of treatment (e.g.: comparing the effects on the number of car accidents before and after the introduction of a new law). Furthermore, statistical techniques such as “difference in difference”, “propensity score matching”, “discontinuity regression” and “synthetic control” also seek to identify the causal effects of treatments.

sample would not necessarily eclipse the relation between the independent and dependent variables, since this kind of analysis relies on the observation of causal processes, which abide to logic different from that of linear regression. Nevertheless, truncated samples can overestimate certain cases, meaning that it is still of the utmost importance for qualitative researchers to carefully select their observations. Following the same logic, the authors defend the validity of within-case studies, even when they are faced with the issue of no variance on the dependent variable, as this is a bigger problem for researches that end on some sort of regression.

Choosing cases based on the value of variables is an important issue, whether the studies have a more probabilistic approach (traditional of quantitative, population-oriented studies) or one concerned with necessary and sufficient causes (more common to qualitative, variable-oriented studies). Choosing cases based on the dependent variable leads to identifying the necessary causes – for when one happens, the other will always follow – but in order to identify sufficient causes variance of the dependent variable should be admitted. Pérez-Liñán (2010) describes how this process, where the methods of agreement and difference of J. S. Mill follow each other, can be complemented by the use of Qualitative Comparative Analysis (QCA) in order to reach a number of causal configurations that might explain the occurrence of a certain phenomenon.¹¹

A second type of problematic bias is that of omitted variables, because they might nullify the causal relations proposed by the research design. If one states that X causes Y, but fails to understand that there is another variable Z intervening on that relation, for instance by overdetermining the values of X, or having a simultaneous impact on both X and Y, the relation identified between X and Y will, in fact, be spurious. To sum it up, if the phenomenon studied is better described by a group of variables and any of them is left out of the equation, the effect of the variables actually used in the model will either be overestimated or underestimated (WOOLDRIDGE, 2013, page 88).

2 INFERENCE

Inference can be understood, in short, as reaching conclusions about non-observed phenomena by using information from previously observed phenomena (KKV, 1994). When making use of inference, researchers aim towards the generalization of their findings, because they are trying to broaden the patterns perceived on one specific case to the rest of the universe. Thus defined, it can be understood that inference is one of the main objectives of scientific research in the eyes of various methodologists. This section will address how some of these authors (1) justify this move towards generalization and deal with the problems regarding ensuring comparability between cases; and as a follow up we will analyze (2) the two different types of inference proposed by KKV and (3) recommendations by the literature on how to improve inferences.

2.1 Specifying x Generalizing

Przeworski and Teune (1970, page 4) argue that the main objective of social research is to “lead to general affirmations about social phenomena”. Consequently, it is accepted that the *explanadum* of Social Sciences, human or social behavior, can be explained by a reduced number of *explanans*: general laws established via observation.

The search for general laws is a goal proper to Natural Sciences, for example, Physics, which searches for patterns and covering laws that take place despite situational contingencies (ALMOND & GENCO, 1977). That means that the presence of the causes must lead to the predicted result, with zero or close to zero unexplained variation. It is controversial to transport this endeavor into Social Sciences, given that the latter originated from Philosophy – specially from Philosophy of History – and, therefore, have a tendency to focus on exceptions or particularities, instead of on

¹¹ See Freitas and Neto (2016) for a didactic introduction to the method.

universal tendencies. To put it into technical terms, the discussion that cuts across the Social Sciences is whether their vocation is nomothetic, with a focus on proposing general laws, or idiographic, focusing on contingencies¹².

Almond and Genco (1977) bring back a metaphor proposed by Popper that captures this dichotomy well. There are phenomena which are as steady as clocks, behaving in a determined and predictable way, while others are as chaotic as clouds, undetermined and unpredictable. The focus on finding general laws leads the social scientist to prefer “clocks” to “clouds”, which, according to Almond and Genco, is problematic, because human interaction and social systems seldom display mechanic regularity and, unlike molecules in a laboratory, will be moved by elements such as will, objectives, strategies and beliefs. This is why it is important to think about the social world relaxing expectations and thus accommodating anomalies and intentions.

Przeworski and Teune (1970) admit that, in the Social Sciences, propositions that aim to explain a phenomenon universally are seen as unreasonable. Their attention does not fall upon the question of anomalies, but rather on diversity. That is, generalized rules might be inadequate, because they ignore the variety of forms a phenomenon can take across different social systems. For example, certain behavior might be considered “corruption” in one country, but be acceptable in others.

The categories used to propose a law, observed in a specific context, cannot always be automatically transposed into other scenarios. The understanding that underlies this recommendation is that social systems are different amongst themselves to the point that, even if a certain phenomenon is given the same name (e.g.: “protest”, “democracy”), its materialization in different environments will be accompanied by so many local idiosyncrasies that the same concept may no longer be useful to describe the same thing.

In regard to this issue, Przeworski and Teune (1970) take the stance that comparative research can indeed yield nomothetic affirmations. In other words, instead of being content with stating that certain phenomena will be observed in specific systems, that is, using the systems’ proper nouns (e.g.: democracy in India, Canada, Nigeria, etc) as predictors, it is possible to replace those names by measurable variables.

Seeking generalization is not putting abstraction before history. Rather, historically situated statements might be generalized, for they cover with specific names (e.g.: “Germany in between wars” or “Africa during the 1960s”) a series of potential causal factors. These may be transferred across contexts with the goal of reaching a more general theory, leaving the unique and non-replicable factors as “residue” of the theoretical explanation. Because of that, Przeworski and Teune¹³ argue that the link between historical observations and generalized theories lies in the replacement of proper nouns by relevant variables (when it is believed that a certain class of social events might be broadened beyond that specific system) and that this is the end goal of comparative research. To the authors, the success in comparing the same element across different situations depends on the criteria of reliability and validity, because in order to replace proper nouns by analytical variables, it is necessary to ensure that these variables do, in fact, capture the phenomenon of interest.

It can be said that two things will be comparable on those attributes that they share: if they do not share any, then they are incomparable; if they share all attributes, then they are the same thing – in both cases comparison is useless. For all the situations in between, comparison is doable. Sartori (1991), however, warns researchers of the risk they face when, aiming to include a higher number of cases in their analysis, they end up stretching the concept to the point where it is emptied of any semantic accuracy, making it operationally useless¹⁴.

12 Przeworski and Teune, *idem*.

13 *Idem*

14 Collier and Mahon (1993) revisited the issue of conceptual stretching and proposed some additional tools to enable comparison between phenomena, such

2.2 Descriptive inferences and causal inferences

KKV (1994) state that the goal of scientific research is inference based on empirical information. The acquired data underpins inferences about things beyond what is directly observed.

The authors recognize two types of inference. The first one is descriptive inference, where observations are used to learn about unobserved facts. Describing is also inferring if we consider that categories and typologies developed from a restricted number of phenomena might possibly encompass other phenomena which have not yet been observed.

The second type of inference is causal inference. According to the authors, this is desirable as often as possible to complete the frame already offered by description. Therefore, we can understand that, to KKV, description precedes causal explanation.

The authors propose three criteria to judge how valid an inference is. Firstly, inferences must be *unbiased*. There is bias whenever replication leads to systematic error, which, at the limit, pushes the results away from the original average¹⁵. Secondly, in order to differentiate between estimators that are equally unbiased, it is appropriate to choose based on *efficiency*, meaning, when there is little variance. Variance will be smaller in studies with a higher number of observations, which is a consequence of the third criterion, *consistency*, which establishes that the bigger the number of observations, the smaller the variance.

2.3 Improving inferences: indetermination, number of observations and multicollinearity

KKV (1994) warn about the issues surrounding undetermined research designs. These happen in two instances: when one tries to achieve more inferences than the number of available observations allows and when there is multicollinearity.

The first issue refers to a dilemma brought up by Lijphart (1971, page 686): “too many variables, small n”. When there are fewer observations, it is hard to sustain that a proposed causal relation happens for one reason rather than for another. Therefore, the determination of a research design refers to the proportion between variables and observations. Hence, KKV and other authors recommend increasing the number of observations as a way to solve the indetermination problem¹⁶. KKV suggest that, when increasing the size of N isn’t possible, the focus of the research should switch from trying to find the cause of effects to focusing on the effects of the causes, or even that the researcher tries to find more observations on a different level of analysis (e.g.: finding data on the state level rather than on the national level). Repeating observations across time (creating so called panel data and time series cross section datasets) is another strategy to increase N.

Munck (2004) questions these solutions by saying that they are not always available: certain political phenomena are not plentiful in human history. Revolutions or wars, for example, happen rarely. Besides, trying to put the largest amount of cases under the same label increases the likelihood of running into conceptual stretching or causal heterogeneity, that is, situations in which the phenomenon is best described by distinct causes. The underlying goal of KKV is to search the largest inferential leverage; however, Munck (2004) argues that, for qualitative research, this would come more from intense knowledge by the researcher about the case and context than from the inclusion of more cases.

as the concept of “family resemblance” and “radial categories”.

¹⁵ More specially, bias happens when the collected samples of a population generate estimates that are systematically different from the values of the real parameters, as distributed amongst the population (KENNEDY, 2003).

¹⁶ See also Sartori (1991).

Collier, Brady and Seawright (2004), in turn, also question the inclusion of more cases as the best way to enhance inference, and underscore that contextual knowledge is equally important for inferential validity. They propose a distinction between data-set observations (DSO) and causal-process observations (CPO). The first are observations in the traditional statistic sense (i.e.: points in a scatterplot) and serve as the foundation for correlational causal inferences.¹⁷ As for CPO, they are defined as insights or information that offer knowledge on the context and mechanisms, contributing to the understanding of the causal process, albeit they might not be entered in a table as new cases or values of an explanatory variable. Increasing the number of DSOs meets the traditional requirement of increasing the size of N and consequently increasing the degrees of freedom, increasing CPOs does not affect the size of N nor the amount of degrees of freedom¹⁸; therefore, the authors argue that both processes aid inference, though through different approaches.

The second threat to inference is multicollinearity. This happens when there is a high degree of correlation between the independent variables¹⁹. The presence of this issue might seriously impair the inferential leverage of the model used by the researcher.

The biggest issue faced by models with multicollinearity problems is the increase on the magnitude of the variance of the estimated parameters. This happens because the presence of high level of correlation between independent variables makes it impossible to estimate with precision the effect of each variable on the dependent variable [...] the bigger the correlation between the independent variables, the least information will be available to estimate the coefficients associated with the explicative variables. (FIGUEIREDO FILHO et al., 2011, pages 57-58)²⁰

According to KKV (1994), randomized selection of observations is one of the mechanism in the quantitative tradition to avoid this type of indetermination, but the authors acknowledge that this process might not be the most practical for certain types of phenomena analyzed by the qualitative tradition. Hence, the authors suggest collecting more observations (preferably cases where the explicative variables are not strongly correlated), be it on the same level of analysis or on a different one. Alternatively, the number of variables used might be limited – at least up until the point where it is possible to carry on without incurring on omitted variables bias. A different technique is compressing some variables from the model into one composite index, via dimensional reduction techniques (factorial analysis, for example) (FIGUEREIDO FILHO, SILVA, DOMINGO, 2015).

Summing it up, the researcher will often come across with the situation depicted on Figure 4. The horizontal axis represents the number of independent variables of the model and the vertical axis the number of cases. Studies with few independent variables (left side) will be more efficient, but incur on a higher risk of bias, whereas those on the right side will be less biased, but with a higher chance of being less efficient, as the proliferation of variables reduces degrees of freedom and increases the likelihood of multicollinearity. As for the N, we have seen that KKV considered studies with a high amount of observations as more consistent (smaller variation). However, an increase in observations can

17 Educational tool to visualize correlation and data points: <<http://guessthecorrelation.com/>>

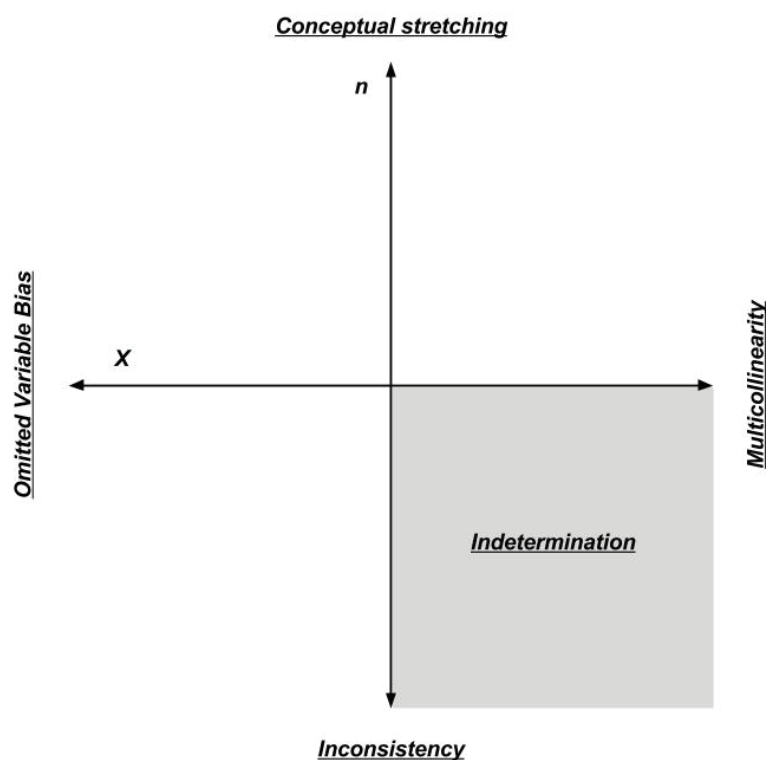
18 The number of degrees of freedom is defined as the total number of observations minus the number of parameters one aims to estimate from them (WOOLDRIDGE, 2013, page 100).

19 It is important to point out that, given the nature of the data used in Social Sciences, there will always be some correlation between variables, meaning that this warning is against high levels or perfect correlation (WOOLDRIDGE, 2013, page 84).

20 In the original: “A maior dificuldade de modelos com problemas de multicolinearidade é o aumento da magnitude da variância dos parâmetros estimados. Isso porque a presença de altos níveis de correlação entre as variáveis independentes impossibilita estimar, com precisão, o efeito de cada variável sobre a variável dependente [...] quanto maior a correlação entre as variáveis independentes, menos informação estará disponível para estimar os coeficientes associados às variáveis explicativas.”

lead to conceptual stretching, according to the type of phenomenon. Indetermination results from the proportion between the mobilized variables and the collect observations. Because of that, studies with too many variables and a small amount of observations (gray area) suffer from indetermination. To sum it up, there is a tradeoff between precision and bias (regarding the variables) and between consistency and conceptual accuracy.

Figure 4: Tradeoff when choosing the number of variables and observations



Source: Elaborated by the author

3 CAUSALITY

The definition of causality used by KKV is similar to Rubin's model (1974 as cited by KKV, 1994, page 82), in which "causal effect is the difference between the systematic component of observations made when the explanatory variable takes one value and the systematic component of comparable observations when the explanatory variable takes on another value". This definition shows that the authors take on a counterfactual understanding of causation – one of the four we will discuss on this paper.

Estimating causal effects is based on some assumptions that allow us to carry this reasoning forward despite the "fundamental problem of inference", which is that it is impossible to see a phenomenon simultaneously happening and not happening in the same unit (REZENDE, 2017). The first assumption is that of unit homogeneity. Two units will be homogeneous when their expected values in the dependent variables are the same when the independent variables takes on a certain value. As it is not always possible to establish this equivalence in the real world, assuming a causal effect that is constant for different units could be enough. The second assumption is conditional independence,

which means that the values taken on by Y do not affect the values of X. The violation of this assumption is highlighted as one of the causes of endogeneity.

3.1 One or many causal logics?

In Political Science and International Relations, a “cause” does not mean the same to everyone. Is there only one rationale through which X is understood as causing Y? Or even, what does it mean to say that X causes Y? Depending on the author, different typologies are proposed to class causal logics²¹. For the purposes of this work, we will highlight the following possibilities:

1) Covariation/Correlational: there is causality when X and Y jointly vary. It is the kind of explanation that validates, for example, methods of simultaneous variation, such as J. S. Mill’s method of difference (named the Most Similar Systems by Przeworski and Teune, 1970). If there is a strong association between theoretically oriented variables, it is expected that there will be a causal relation between them – although correlation is not the same as causation, as correlations can be spurious.

2) Mechanisms: it is impossible to speak of a causal relationship simply by measuring the values of two variables and ignoring what happens between them. There is no “action at a distance” (DAHL, 1957, page 204). It is necessary to tackle the mechanisms that logically connect a cause to its consequence. Methods such as process tracing and historical analysis are concerned with bringing these links to light.

3) Experimental: causation is linked to manipulation, that is, if the researcher can manipulate the variables, control for biases, and observe effects, then he/she can state with confidence that there is causation.

4) Counterfactual: answering the question “how would this case behave if this specific element were removed?” is one of the ways to establish a causal relation. However, this kind of rationale falls into the fundamental problem of inference. That is why comparative research is relevant: it aims to be a sort of control in which one tries to measure the effect of the same variables in different contexts (SARTORI, 1991).

For some authors, wary of the implications of such causal pluralism on the scientific status of Political Science, this syncretism must be brought to a single faith. Gerring (2005) states that, although the pluralist approach to causality has contributed to add some complexity to “naive positivism”, it is also problematic for exacerbating the difference between perspectives which are, in fact, complementary. According to the author, this segregation compromises the goal of accumulating scientific knowledge. He proposes a minimum definition, according to which “cause” is one (or various) “events or conditions that raise the probability of some outcome occurring (under *ceteris paribus* conditions). X may be considered a cause of Y if (and only if) it raises the probability of Y.” (GERRING, 2005, page 169).

However, it is interesting to point out that even such monist projects give birth to separate doctrines. Mahoney (2008), in turn, sees problems in seeking unity via probabilistic notions. This is because we can in the methodological literature a distinction between case-oriented research and population-oriented research (RAGIN, 2004; COLLIER, BRADY SEAWRIGHT, 2004; MAHONEY, 2008). While the former seeks to explain what causes specific results in specific contexts, the latter wants to identify mean or marginal effects on populations. This has consequences for the conception of causation, because the goal is, for one, explaining specific results and, for the other, statistic probabilities. Mahoney (2008) understands that, in order to conceive causation in a single way, it must be admitted that a cause on the individual level is logically previous to a cause on the population level. Because of that, the minimum

21 See Gerring (2005) for a list.

definition “cause = an increase in probability”, as Gerring (2005) proposes, cannot be adopted, given that it makes no sense to talk about probability when $N = 1$. Therefore, Mahoney (2008) presents a definition of causality on the individual level in terms of necessary/sufficient conditions.

Beyond the traditional combinations of analysis that deal with necessary/sufficient causes (1. Necessary and sufficient cause; 2. Necessary but insufficient cause; 3. Sufficient cause but unnecessary cause) the author proposes two more (taken from MACKIE 1965 as cited in MAHONEY, 2008): (4) INUS causes (Insufficient but Necessary part of an Unnecessary but Sufficient condition), for which there is no single necessary or sufficient cause. The individual cause is, in fact, an insufficient but necessary part of an arrangement which is, in turn, unnecessary but sufficient.

For example, take:

$$Y = (A \ \& \ B) \mid (C \ \& \ D).$$

Putting it into formal terms, “A” is insufficient because it needs to be associated with “B” (“&” represents the logical “and”), but this pair is not needed, because if the pair “C & D” happens, the phenomenon Y will happen anyways (“|” represents the logical “or”). As for the (5) SUIN causes (Sufficient but Unnecessary part of an Insufficient but Necessary context), they are those where the individual cause is the sufficient but unnecessary part of a factor, and this factor is, in turn, the insufficient and necessary cause of a phenomenon. In formal terms, we have:

$$Y = X \ \& \ Z; \text{ where: } Z = A \mid B.$$

“A” is sufficient but not necessary to create “Z”, and “Z” is insufficient but necessary to create “Y”. Therefore, “A” (and also “B”) is to “Y” a SUIN cause.

In order to unify this deterministic typology with a probabilistic understanding of causation, the author proposes that causes happen on the population level following the INUS rationale, as the growing heterogeneity on population level that makes the (5) SUIN, (1) necessary and sufficient, and (2) necessary and insufficient causes rarer.

One of the consequences of the INUS causes is equifinality, that is, different groups of variables lead to the same result (the pairs “A & B” or “C&D” on the first example). This characteristic is tacitly implied in linear models, because these usually have many variables and there is a virtually infinite number of combinations of their values that will yield the same numerical score on Y. This means that no cause on the model is necessary, but different values on the variables (associated with the error term) are sufficient to produce the result.²²

3.2 Ontological assumptions: Probability x Determinism

One of the main divisions between schools of thought in Political Science and International Relations has to do with how the researcher understands reality²³. The world is either probabilistic, meaning that even if all the possible independent variables are summoned to explain a certain phenomenon, it is not possible to say exactly what will happen,

22 Perfecting this model later on, Mahoney, Kimball and Koivu (2009) seek to solve a challenge of non-statistic research: if multiple causes are interacting to create a phenomenon, how do we define which one is the most important? While linear regressions can answer this question by looking into the value of the adjusted coefficients, for instance, which will show which variables have the biggest effects on Y (see FIGUEIREDO FILHO et al, 2011, page 49), for historical analysis they propose the “method of sequence elaboration”, where the identity between the independent and the dependent variables might be used as criterion for relevance.

23 The division presented by Almond (1988) between ideology (left/right) and methodology (hard/soft) also takes into consideration whether the researcher believes it is possible to create objective scientific knowledge. This political dimension of the discipline is not addressed in our article.

because chance will always play a part; or the world is deterministic: things either happen or do not happen, *tout court*, and thus any apparent randomness is because there is still some independent variable that was not included.

KKV (1994) favor the probabilistic approach. They recommend, therefore, that one differentiates between systemic elements (i.e.: which must have their causal effect regularly explained by the theory) and non-systemic ones (i.e.: effects that are not covered by the theory; noise), based on a priori criteria about predictability and randomization. Because of that, although they recognize the Popperian principle of falsifiability as extremely important to scientific endeavor, they relativize it. They understand that the falsification of a certain theory by new evidence is rather an indicator of the theory's limitation, instead of a verdict condemning it to rejection.²⁴

Ragin (2004, page 124) considers that case-oriented research and variable-oriented research are “a different mode of inquiry with different operating assumptions”. Specifically, they have different practical concerns: (1) case constitution: in case-oriented research, cases are not seen as homogeneous observations from a pool, but as complex configurations of events. (2) Defining negative cases: case studies are often choosing only on the positive values of their dependent variables, which is problematic for generalization. According to Ragin (2004), concerns about generalization are already taken into consideration during the case selection process, when concerns about sufficient and necessary conditions are implicitly mobilized. (3) Analysis of conjectural and multiple causes: case-oriented researches see causation as dependent on conjectures and therefore they are not useful to measure the independent effect X has on Y, as variable-oriented research can.

Finally, (4) the treatment of deviant cases: while variable-oriented research tends to consider these as residues or noise, case-oriented research will seek to explain them substantively, which means that each approach has a different understanding of what should be done with the “error” of a theoretical model.

These different conceptions affect the validity of some types of research design. For a deterministic approach, that considers the causal universe as invariant, a single case study might be enough to overthrow a theory (ROGOWSKI, 2004). A probabilistic approach would, however, treat the findings of a case study with a little bit more parsimony.

4 CONCLUSION

The current article has provided a didactic introduction to the main discussions surrounding research design in Political Science and International Relations. As it has been shown, our discipline has accumulated a robust framework of recommendations about how to better one's research design, inference and claims of causality. Some elements are common ground amongst most authors, such as the importance of choosing relevant research question and the concern with conceptual validity and data reliability. Other points, however, are less consensual.

In particular, it was possible to show that the differences point towards two different polarized traditions. On the one hand, we find the large n, quantitative research, which is concerned with populations/variables and with measuring the effects of causes and marginal impact, holding a probabilistic world view; on the other hand, we find small n, qualitative research, which is concerned with cases and the cause of effects, holding a deterministic world view. We could further add that the former prizes generalization while the latter is focused on explaining specific events in detail, but one of the points of contention is precisely whether or not the latter school of thought can be used to achieve generalizable hypotheses. While some authors see these fields as complementing one another, others see differences which point to a completely different rationale.

²⁴ An example of that would be rational choice theory, which, even with growing evidence against some its tenets, is not considered obsolete or disposable, because it still explains a good portion of human behavior.

Finally, we believe it is important to highlight the emphasis given by KKV in anchoring generalization on observations: this is what makes science different from speculation. However, *political* scientists have an interesting subject matter at hand. Not all phenomena that are relevant for Political Science and International Relations are high frequency “observation generators”. Some do have a high and regular amount of observations (e.g.: elections, financial flows, commercial deals, census, legislation, etc), while can only be seen from afar, rarely and darkly (legitimacy, ideology, bargaining, coercion). When it comes to politics, depending on the area, the facts that come to light are only a fraction of those that remain hidden. Therefore, it is beneficial, on the one hand, to be humble about how far our observations can go when it comes to political processes; and, on the other hand, avoid the trap of considering certain studies more sophisticated simply because their object of choice is one that yields more and better data.

What this exposition makes clear is that scientific research demands tradeoffs and that it is up to the researchers to be very aware of their questions, topic of choice, the possibilities and limitations that empirical reality imposes on this endeavor and, based on that, opt for the right toolbox for the project. The inferential frontier will always be set by factual reality and what it gives us to work with.

With this article, we hope to have introduced to undergraduate students the main discussions about research design in Political Science and International Relations nowadays. This knowledge is a fundamental starting point for both interacting critically with the literature produced by their peers and for the conscious decision on how to deal with a research question.

BIBLIOGRAPHY

- ALMOND, Gabriel A. (1988). "Separate tables: schools and sects in political science" *PS: Political Science and Politics*. Vol. 21, No. 4: 828-842
- ALMOND, Gabriel A.; GENCO, Stephen J. (1977) "Clouds, clocks and the study of politics". *World Politics*. Vol. 29, No. 4: 489-522
- BARBOZA, Danilo P.; GODOY, Samuel R. (2014) "Superando o 'calcanhar metodológico'? Mapeamento e evolução recente da formação em métodos de pesquisa na pós-graduação em Ciência Política no Brasil". *Trabalho apresentado no IV Seminário Discente da Pós-Graduação em Ciência Política da USP*. 07 a 11 de abril de 2014.
- BLALOCK Jr, H. M. (1961). "Theory, Measurement, and Replication in the Social Sciences". *American Journal of Sociology*, Vol. 66, No. 4: 342-347
- BRADY, Henry; COLLIER, David. (2004). *Rethinking social inquiry: diverse tools, shared standards*. Maryland: Rowman & Littlefield Publishers.
- COLLIER, David; BRADY, Henry; SEAWRIGHT, Jason. (2004), "Sources of leverage in causal inference: towards an alternative view of methodology" in Henry Brady; David Collier (eds). *Rethinking social inquiry: diverse tools, shared standards*. Maryland: Rowman & Littlefield Publishers, pp. 229-266
- COLLIER, David; MAHON, James E. (1993). "Conceptual "stretching" revisited: adapting categories in comparative analysis." *American Political Science Review*, Vol. 87, No. 4: 845-855
- DAHL, Robert A. (1957). "The Concept of Power". *Behavioral Science*. Vol. 2, No. 3: 201-215
- FIGUEIREDO FILHO, Dalso; SILVA, Lucas; DOMINGOS, Amanda. (2015). "O que é e como superar a multicolinearidade? Um guia para a ciência política." *Conexão Política*. Vol. 4, No. 2: 95-104
- FIGUEIREDO FILHO, Dalson, et al. (2011), "O Que Fazer e o Que Não Fazer Com a Regressão: pressupostos e aplicações do modelo linear de Mínimos Quadrados Ordinários (MQO)." *Revista Política Hoje*, Vol. 20, No. 1
- FREITAS, Vitor S.; NETO, Fernando B. (2016). "Qualitative Comparative Analysis (QCA): usos e aplicações do método". *Revista Política Hoje* Vol. 24, No. 2: 103-118
- GEDDES, Barbara. (1990), "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics." *Political Analysis*. Vol. 2, No. 1:131-150
- GERRING, John. (2004), "What is a case study and what it is good for?" *American Political Science Review*, Vol. 98, No. 2: 341-354.
- GERRING, John. (2005), "Causation: a unified framework for the social sciences." *Journal of Theoretical Politics*, Vol. 17, No. 2: 163-198
- KENNEDY, Peter. (2003). *A guide to econometrics*. Cambridge: MIT Press.

- KEOHANE, Robert O.; NYE, Joseph N. (2001). *Power and interdependence*. New York: Pearson.
- KING, Gary. (1995), "Replication, replication." *PS: Political Science & Politics*. Vol. 28, No. 3: 443-499.
- KING, Gary. (2015) [1995]. "Replicação, replicação". *Revista Eletrônica de Ciência Política*, Vol. 6, No. 2: 382-401
- KING, Gary; KEOHANE, Robert O.; VERBA, Sidney. (1994), *Designing Social Inquiry: scientific inference in qualitative research*. Princeton: Princeton University Press.
- LIJPHART, Arend. (1971), "Comparative politics and the comparative method." *American Political Science Review*, Vol. 65, No. 3: 682-93
- MAHONEY, James. (2008), "Towards a unified theory of causality." *Comparative Political Studies*. Vol. 41, No. 4-5: 412-436
- MAHONEY, James; KIMBALL, Erin; KOIVU, Kendra L. (2009), "The Logic of Historical Explanation in the Social Sciences." *Comparative Political Studies*, Vol. 42, No. 1: 114-146
- MEDEIROS, Marcelo de Almeida et al. (2016). "What does the field of International Relations look like in South America?" *Rev. bras. polít. int.*, Vol. 59, No. 1, e004, 2016
- MERTON, Robert K. (1942). *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press.
- MILLER, Bernhard. (2011), "Making measures capture concepts: tools for securing correspondence between theoretical ideas and observations." in Thomas Gschwend; Frank Schimmelfennig (orgs.). *Research Design in Political Science: how to practice what they preach*. London: Palgrave Macmillan, pp.83-102
- MUNCK, Gerardo. (2004), "Tools for qualitative research." in Henry Brady; David Collier. *Rethinking social inquiry: diverse tools, shared standards*. Maryland: Rowman & Littlefield Publishers, pp. 105-122
- NEIVA, Pedro. (2015). "Revisitando o calcanhar de Aquiles metodológico das ciências sociais no Brasil". *Sociologia, Problemas e Práticas*. No. 49: 65-83
- PEARSON, Karl. (1900) *The Grammar of Science*. Londres: Adam and Black. Disponível em: < <https://archive.org/details/details/grammarofscience00pearuoft> > Data de acesso: 12/08/2017
- PÉREZ-LIÑÁN, Aníbal. (2010), "El método comparativo y el análisis de configuraciones causales." *Revista Latinoamericana de Política Comparada*. Vol. No. 3: 125-148
- POPPER, Karl. (2002) [1935]. *The logic of scientific discovery*. Londres: Routledge.
- PRZEWORSKI, Adam; TEUNE, Harry. (1970), *The Logic of Comparative Social Inquiry*. New York: Wiley-Interscience.

- RAGIN, Charles. (2004), "Turning the tables: how case-oriented research challenges variable-oriented research." in Henry Brady; David Collier (eds.). *Rethinking social inquiry: diverse tools, shared standards*. Maryland: Rowman & Littlefield Publishers, 123-138
- REZENDE, Flávio da Cunha. (2017). "O Pluralismo Inferencial na Ciência Política Pós-KKV (2005-2015): Argumento e Evidências". *Revista Política Hoje*. Vol. 26, No. 1: 241-278
- ROGOWSKI, Ronald. (2004), "How inference in the social (but not the physical sciences) neglects theoretical anomaly". In Henry Brady; David Collier. *Rethinking social inquiry: diverse tools, shared standards*. Maryland: Rowman & Littlefield Publishers, pp. 75-84
- SARTORI, Giovanni. (1991), "Comparing and miscomparing." *Journal of Theoretical Politics*. Vol. 3, No. 3: 243-257
- SILVA, Glauco Peres da. (2015). "Desafios ontológicos e epistemológicos para os métodos mistos na ciência política". *Rev. bras. Ci. Soc.* Vol.30, No.88: 115-128
- SOARES, Gláucio. (2005), "O calcanhar metodológico da ciência política no Brasil." *Sociologia, problemas e práticas*, No. 48: 27-52
- TARROW, Sidney. (2005), "Bridging the quantitative-qualitative divide." in Henry Brady; David Collier. *Rethinking social inquiry: diverse tools, shared standards*. Maryland: Rowman & Littlefield Publishers, pp.171-180
- WOOLDRIDGE, Jeffrey M. (2013). *Introductory Econometrics: a modern approach*. 5th edition. Ohio: South-Western.