Regular Article

Journal of Peace Research

Bias mitigation in empirical peace and conflict studies: A short primer on posttreatment variables

Journal of Peace Research
1–15

DOI: 10.1177/00223433221145531
journals.sagepub.com/home/jpr
SAGE

Christoph Dworschak

Department of Politics, University of York

Abstract

Posttreatment variables are covariates that are preceded by the main explanatory variable. Their inclusion in a statistical model does not 'control' for their influence on the relationship of interest, and it does not substitute for a mediation analysis. Likewise, a coefficient estimate of an appropriate 'control variable' cannot be interpreted as a causal effect estimate. While these facts are well-established in various fields across the social sciences, their recognition in the field of peace and conflict studies is more limited. Originally collected data on recent publications from leading peace and conflict journals reveal that a large majority of evaluated articles condition on posttreatment variables, demonstrating how a review of these fallacies can help to substantially improve future research on peace and conflict. Drawing on a broad set of literature and using graphical approaches, I offer an intuitive explanation of the logic of posttreatment variables and clarify common misconceptions. Building on recent developments in methodology and software, and by deriving conditions for bounding using analytical bias expressions, I discuss avenues for dealing with posttreatment variables in observational studies. The article concludes with a discussion of implications for applied research.

Keywords

causal inference, model specification, peace and conflict, political analysis, posttreatment bias

Introduction

Which variables should researchers *not* condition on (not 'control for') in empirical research on peace and conflict? Research design and variable selection are areas in which there are no easy answers available. While the computation of a regression is usually just one click away, which covariates to include in that regression no computer can tell (King, Keohane & Verba, 1994).[1] Therefore, questions of designing research and selecting variables have been studied abundantly. This article attempts to raise renewed awareness to the challenge of variable selection and discusses avenues to address common issues that are particularly relevant to applied research. In doing so, emphasis is given to straightforward and accessible explanations rather than to statistical depth. The target audience of this article are empirical peace and conflict researchers.

Most quantitative research on peace and conflict seeks to approximate causal claims using observational data. While observational research designs can never match the gold standard of design-based inference, the use of appropriate statistical methods, availability of high-quality data and careful model design can go a long way. The practice of using 'control variables,' that is, of conditioning on covariates to partial out the effect of an

[1] This and other generalizing statements in this article assume a deductive quantitative research design seeking to approximate causal claims in the context of an observational null-hypothesis significance testing framework. While the lessons drawn here equally apply to other empirical approaches, including experiments and qualitative comparison, I adopt a more targeted language due to scope constraints and to improve accessibility.

Corresponding author:
christoph.dworschak@york.ac.uk

explanatory variable of interest, attests to scholars' effort to transcend claims of mere correlation (King, Keohane & Verba, 1994).[2]

In support of this effort, I argue that empirical research in the field of peace and conflict studies does not pay enough attention to the question of *causal sequence*. The causal ordering of variables matters for the estimation of causal effects. One well-known example of the importance of sequence is the topic of 'reverse causality:' when estimating the effect of an explanatory variable of interest (the 'treatment;' X) on a variable to be explained ('outcome;' Y), the estimate may be distorted by a reverse effect, that is, not only does X influence Y, but also does Y influence X. An example is the effect of democratization on economic prosperity, and vice versa. The importance of reverse causality is well-established and commonly considered in the curriculum, reviewer comments, conference discussions and editor reports. This article focuses on an issue of similar importance and related to causal effect direction, but which receives much less attention: treatment effect estimates are sensitive to the causal sequence of the covariates included for conditioning (also called 'confounders' or 'control variables;' Z). If the treatment variable causally precedes a covariate, this covariate is called a 'posttreatment variable' and its inclusion in the analysis biases the treatment's total effect estimate. This concern over causal direction between the treatment and covariates should receive just as much attention as the question of causal direction between the treatment and the outcome.

Addressing the potential for posttreatment bias can be simple, while ignoring it can substantially distort estimation results – even making the coefficient point in the opposite direction – and lead to an erroneous (failure of) rejection of the null-hypothesis. Using graphical illustrations and drawing on a broad body of methodological literature, I examine core concepts and provide an accessible explanation of why peace and conflict research should care about posttreatment variables. Reviewing classical approaches and recent methodological advances, and by analytically deriving conditions for bounding exercises, I reflect on avenues for how to avoid common sources of bias related to variable selection, including omitted variable bias (OVB), selection bias and overcontrol bias.

This article contributes to the peace and conflict research programme and to social science methodology in several ways. A review of all publications between January 2018 and May 2021 in the *Journal of Peace Research* (*JPR*) and the *Journal of Conflict Resolution* (*JCR*) indicates that 75% of relevant articles may suffer from posttreatment bias and, therefore, may report substantially misleading results. Only a fraction of these studies shows any awareness of this issue. As I will show below, many mistakes can be easily avoided or addressed. Therefore, an accessible explanation of the role of posttreatment variables provides an important opportunity to substantially improve empirical research in the field of peace and conflict. At a more foundational level, the article touches on core empirical concepts and reiterates the role of covariates in multiple regression. The extent to which this may seem substantively trivial is exactly what underlines its paramount importance: many years after the works of, for example, Achen (2005) and Clarke (2009, 2005), my review of publications in peace and conflict studies indicates that some of the most basic tenets still find limited application. Including the 'usual set of controls' without considering the bias they may induce, and a standard to interpret covariate coefficients, are among the practices that warrant renewed attention.

While the main goal of this primer is to provide a pedagogical and targeted read on the topic of causal sequence, it also offers a few methodological innovations. First, drawing on a broad set of literature across the fields of political science, psychology, biostatistics and epidemiology, it offers a concise overview of a large number of technical contributions that, directly or indirectly, touch on the topic of posttreatment bias. In doing so, it highlights how the separate literatures on variable selection, reverse causality and mediation analysis intersect and can be applied to conceptualize posttreatment variables. For example, as of this writing there is no dedicated study on the practice of lagging covariates to ameliorate posttreatment bias, but some lessons can be inferred from research on reverse causality. Second, it is the first to offer a systematic assessment of the conditions under which a treatment effect can be bounded by including and excluding a 'proxy control' while allowing for collider-stratification bias. Using analytical bias expressions, I show that the number of scenarios that allow bounding based on effect directions is very limited and

[2] I draw on basic causal inference terminology throughout this article, such as 'treatment' and 'outcome.' Some researchers perceive the use of 'causal language' in the context of observational studies as inappropriate. However, most observational research on peace and conflict is interested in testing causal theories. Using causal language makes the goal of approximating causal inference explicit rather than implicit, and helps to clarify core assumptions underlying these studies. See, for example, Hernán (2018) for an extended discussion on terminology in observational causal inference.

how, among these few scenarios, bias is unevenly distributed.

The article proceeds as follows. First, I review recent publications in *JPR* and *JCR* to illustrate the extent to which posttreatment bias may threaten inference in the contemporary peace and conflict research programme. I then systematically introduce and explain the concept of posttreatment variables and the distinction between confounding and mediation. Third, I outline why conditioning on a posttreatment variable biases estimation in the context of different research objectives, namely the total and direct treatment effects, and how to recover unbiased results in a simple research setup. This is followed by a discussion of the challenge of 'proxy controls' in applied research and potential avenues to address it. These include computational sensitivity analyses, deriving bias directions for bounding in analytical sensitivity analyses, the lagging of covariates and the total effect decomposition approach. The article concludes with a summary of implications and an abridged checklist of key takeaways.

But everybody knows this, right?

Posttreatment bias is a known issue in the social sciences and systematic scholarship on it dates back to Rosenbaum (1984). To understand the extent to which observational research on peace and conflict considers this threat to inference, I collected data on all publications in *JPR* and *JCR* between January 2018 and May 2021. Of all articles that employ a 'standard' regression framework,[3] 75% condition on covariates that may be influenced by the treatment (249 out of 334) in their main analysis.[4] The proportions are visualized in Figure 1.

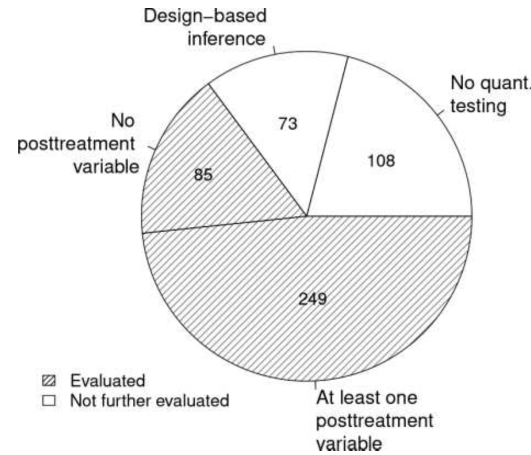

Figure 1. Article sample

Note: Numbers refer to the absolute number of articles corresponding to each slice. Articles without a quantitative hypothesis test or without covariates (design-based inference), represented by the two blank slices, were excluded from further coding.

In other words, three-quarters of reviewed articles may report biased results. The most common offenders are so-called 'standard control variables,' such as gross domestic product per capita and regime type, which are often included in regression models without considering their necessity and appropriateness as covariates. Another common practice among the evaluated studies is to include multiple treatment variables relating to different hypotheses in the same regression model, so that they all condition on each other. As I will show below, the resulting bias can be substantial: including an inappropriate covariate is just as problematic as failing to account for relevant confounders and means risking substantively wrong conclusions and misleading policy recommendations. In other words, using an inadequate 'control variable' is just as problematic as failing to 'control for' an adequate one.

[3] Standard regression framework refers to all articles that conduct a quantitative null-hypothesis significance test with identification based on observables. As mentioned above, design-based inference and qualitative studies can suffer from posttreatment bias just as quantitative inferential approaches that rely on identification based on observables (Montgomery, Nyhan & Torres, 2018). However, data gathering focused solely on the latter type of publications to streamline coding and in line with the emphasis of this article. 'Design-based' is sometimes also used to describe randomization inference as opposed to model-based inference; instead, here it is used as shorthand to refer to experimental and quasi-experimental approaches that do not rely on the conditioning of covariates in the context of a statistical model.

[4] This proportion is consistent with a similar review conducted in Acharya, Blackwell & Sen (2016) of major political science journals. Coding covariates' potential for inducing posttreatment bias requires careful theoretical and empirical consideration of the relationship between each covariate and the respective treatment variable. Therefore, while data are reported at the article level, each covariate was coded at the level of individual hypotheses. Such an assessment ultimately relies on the implications of the theoretical model of the data generating process. Therefore, the coding task required a certain degree of subjective evaluation, which is why the coding process was designed to err on the conservative end (minimizing false positives at the risk of increasing false negatives). For example, all covariates that were lagged to temporally precede the treatment were automatically conceived as being pretreatment, irrespective of the viability of using lagging to this end (see discussion on lagging below). Each hypothesis of each article was coded at least twice, by different coders. At the end of the data project, the coders followed a consensus procedure in which they reconciled coding differences for each hypothesis. Details of the coding process, coded variables, and steps taken to maximize reliability are documented in the codebook.

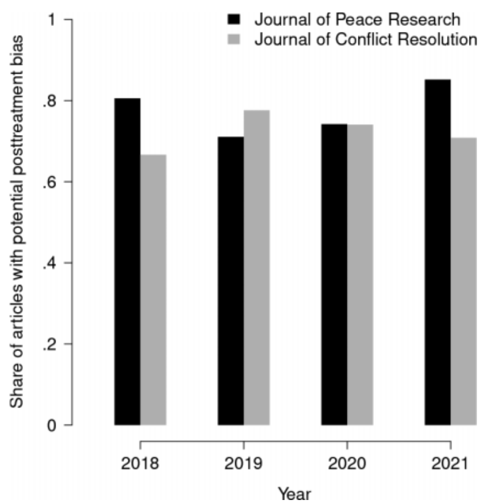

Figure 2. Journal comparison over time
Share of articles among all studies that were evaluated.

As attention to posttreatment bias has increased in recent years, it may be argued that contemporary awareness among peace and conflict researchers is higher than suggested by the pie chart, and that it just takes time for this new generation of research to fully emerge. However, Figure 2 does not support this notion. Comparing the yearly share of studies that include posttreatment variables, there is no sign of improvement over time.

Alternatively, it may be argued that this high share of studies that include posttreatment variables is not due to a lack of awareness, but intentional: as I will show below, deciding on the inclusion of covariates can present a difficult trade-off. Sometimes, choosing to condition on a 'bad control' (Angrist & Pischke, 2009) is a conscious decision in favour of mitigating OVB and accepting the potential of posttreatment bias. Such a decision requires careful consideration of the assumed data generating process and empirical model. To learn the degree to which the high share of inclusion of posttreatment variables is based on conscious decisions, information on manuscripts' operationalization and results discussion were also coded. Out of all 249 articles that include at least one posttreatment variable in their main analysis, only 28 show any awareness of this issue in their discussion.[5] This lack of transparency in the face of possibly large bias that may, in some cases, substantially distort the results is concerning and suggests unawareness among peace and conflict scholars.

This conclusion finds additional support in the fact that 63% of all coded articles include an explicit interpretation of covariate coefficient estimates. As I will discuss below, in most applied cases in which researchers use covariates to minimize OVB in a treatment effect estimate, interpreting the covariates' coefficient estimates is not tenable. Put differently, the widespread norm in peace and conflict research of interpreting 'control variable results' is, at best, futile and, at worst, substantially misleading.

In sum, the topic of 'control variables' requires renewed attention, and a primer on the relevance of their causal sequence is warranted. As the sample of coded articles shows, this is not an issue pertaining to any individual article, but is something that the field of peace and conflict research faces collectively. Research is not produced in a vacuum. The lack of transparency and discussion of this source of bias, as well as the continued norm of interpreting covariate coefficient estimates, raise important questions not only for individual authors, but for colleagues and supervisors who discuss manuscripts at draft stage, and for reviewers and editors who do so at publication stage. For example, conference discussions and peer review tend to centre on the risk of OVB (the notorious 'Have you controlled for . . . ?') while disregarding over-control bias (asking 'How do you justify controlling for . . . ?').[6] For any research project, even if accepting the potential for posttreatment bias was a tenable option amid worse alternatives, such meaningful decisions warrant a transparent discussion. Below, after a short introduction to key concepts, I provide explanations aimed at helping peace and conflict researchers to navigate these choices and to be transparent about the assumptions they require.

Confounder or mediator?

Observational peace and conflict studies that are interested in the estimation of a directional effect of a treatment variable on an outcome variable include additional covariates in their analysis. These covariates are included for the purpose of mitigating confounding. Not conditioning on a confounder means risking OVB. However, not all variables are confounders and qualify to be held constant, and their causal ordering can give important clues on their adequacy (Gelman & Hill, 2007; Pearl,

[5] As with all other coded information, awareness for the potential of posttreatment bias was coded as favourable as possible. For example, lagging all relevant covariates to temporally precede the treatment is, even in absence of any discussion or explicit justification, treated as 'showing awareness.'

[6] See also Achen (2005), Clarke (2009, 2005) and Schrodt (2013) for related discussions.

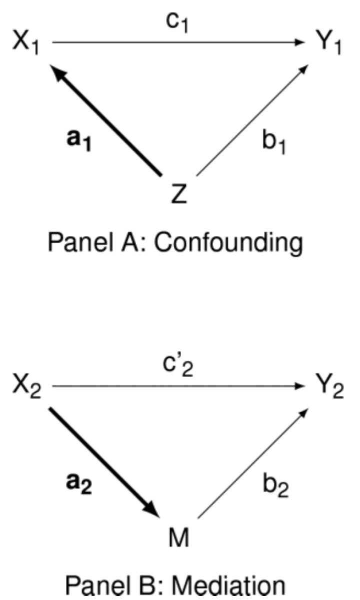

Figure 3. Basic model

Glymour & Jewell, 2016). As a rule of thumb, it is useful to condition on covariates that affect *both* the treatment and the outcome (Imai, 2018: 57–58)[7] – see Panel A in Figure 3 for a schematic. A researcher interested in estimating the total effect of X on Y, here denoted as c_1, would condition on Z in the analysis (that is, include Z as a 'control variable'). Some covariates, however, do not (only) influence X, but are directly or indirectly influenced by X – see Panel B in Figure 3, noting the change in direction of path a. Such variables are causally preceded by the treatment and are therefore called 'posttreatment variables' (M). If they also influence Y (here b_2), then they may constitute a channel through which the effect of X on Y is relayed. In a setup such as the one in Panel B, instead of confounding the relationship between X and Y, they 'mediate' the relationship. In other words, these variables are the mechanisms that link the explanatory variable of interest to the outcome variable.[8]

Most studies are interested in estimating the overall (total) effect of X on Y, which is denoted in Panel A as c_1. In Panel B the total effect c_2 is not shown, because a part of the total effect goes *through* M. Borrowing from mediation analysis language, the effect c'_2 is called the 'direct effect' of X_2 on Y_2, independent of M (Hayes, 2018: 107–108). The effect a_2 b_2 is referred to as the 'indirect effect' of X_2 on Y_2. Taken together, the (natural) direct and indirect effects can be combined as $c_2 = c'_2 + a_2 b_2$, which is the total effect of X_2 on Y_2 in Panel B (Imai, Keele & Tingley, 2010; Pearl, 2014; Hayes, 2018; Pearl, 2014).[9] The total effect c_2 can be estimated by simply regressing Y_2 on X_2, leaving M out of the model. The effect's decomposition is illustrated via simulation in the Online appendix, assuming linearity and using simple ordinary least squares regression.

Figure 4 mirrors Figure 3, exemplifying the variables X, Z, M and Y with actual concepts from peace and conflict studies. This example is loosely based on Wood, Kathman & Gent (2012), though adjusted and simplified for the purposes of illustration. A research hypothesis for this setup may read 'A foreign state's intervention on the side of the rebels during civil war decreases one-sided violence perpetrated by the rebels,' possibly due to a shift in the actors' power balance (cf. Wood, Kathman & Gent 2012). In this case, a foreign state's intervention is the treatment X, and rebel one-sided violence is the outcome Y. Therefore, like most research on peace and conflict, this example hypothesis is geared towards testing the total treatment effect c_1. For estimating this total treatment effect, it is conceivable that rebels' baseline strength (e.g. their relative troop size), before an intervention occurs, influences both their propensity to victimize civilians (outcome), as well as foreign states in their decision on whether to intervene or not (treatment). This makes rebels' strength before the intervention a potential confounder Z, as shown in Panel A of Figure 4.

However, what if rebels' strength is measured *after* the foreign state's intervention has already begun, as shown in Panel B of Figure 4? In this case, rebels' strength is probably influenced by the intervention itself, as shown in Panel B. Therefore, rebels' strength during the intervention is a mediator M: it is a mechanism that relays

[7] As is common for a rule of thumb, this ignores certain caveats for the sake of simplicity: in terms of the potential outcomes framework, we wish to include covariates so that the potential outcomes are conditionally independent of treatment assignment; and in terms of causal graph theory, we wish to condition on a set of covariates as to satisfy the backdoor criterion (cf. Pearl, 2009).

[8] 'Posttreatment variable' is an umbrella term encompassing all variables causally preceded by the treatment. Not all of them influence the outcome (missing b_2) or are directly affected by the treatment (missing a_2). I specifically focus on mediating posttreatment variables as shown in Figure 3's Panel B, because research on peace and conflict seldom conditions on covariates that do not influence the outcome. In fact, while covariates' relationship with the treatment variable is sometimes ignored altogether, special emphasis is usually put on covariates' effect on the outcome. In addition, estimates are potentially distorted even when b_2 is missing due to endogenous selection bias (Elwert & Winship, 2014).

[9] This only works for the natural direct and indirect effects (Imai et al., 2011) and not with, for example, the controlled direct effect (Acharya, Blackwell & Sen, 2016).

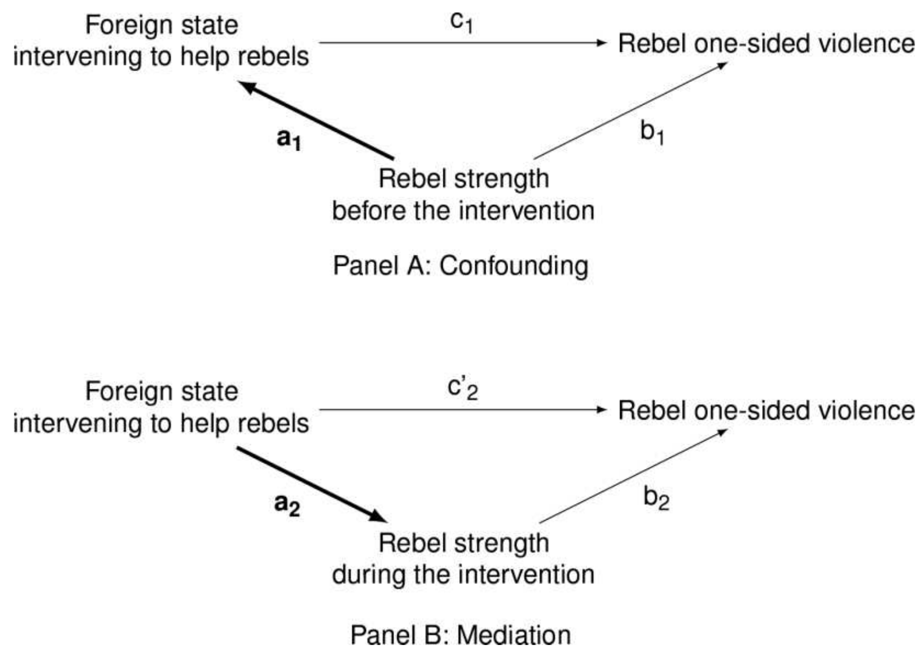

Figure 4. Basic example

part of the effect of the treatment, state intervention, on the outcome, one-sided violence.[10]

In summary, determining whether a variable precedes or succeeds the treatment is necessary to understand whether the variable acts as a pretreatment confounder or a posttreatment mediator. Therefore, when discussing the 'control variables' in a research design, it is vitally important to exercise transparency over the assumed direction between the treatment and each covariate (the direction of path a in Figures 3 and 4). The following section details why this distinction is important for unbiased estimation.

Bias by (not) conditioning and what to do about it

Conditioning on a posttreatment variable M can be problematic.[11] First, and intuitively, conditioning on a posttreatment variable means to partial out a part of the treatment effect itself. The mediating variable acts as a mechanism that relays a part of the effect of X on Y, which is why the arrows in Figure 3's Panel B indicate that part of the effect of X on Y 'flows through' M. Conditioning on it means to exclude a portion of (i.e. biasing) the total treatment effect (Gelman & Hill, 2007; Pearl, Glymour & Jewell, 2016; Cinelli, Forney & Pearl, 2020). Referring back to the example in Figure 4's Panel B, rebels' strength is a mechanism that links a foreign state's intervention to rebel-perpetrated one-sided violence. In this artificial scenario, conditioning on rebels' strength means to partial out that mechanism, thus biasing the total effect estimate of foreign intervention on one-sided violence, and getting an incorrect test result for the hypothesis above. This kind of bias has the intuitive name of 'over-control bias' or 'posttreatment bias' (Elwert & Winship, 2014). The bias can go in either direction, either inflating or attenuating the coefficient estimate. Under ideal circumstances, the effect researchers are left with is the direct effect, c'_2, instead of the total effect c_2, as illustrated above and in the Online appendix Table A.1. However, for reasons discussed in the following paragraph, even this is rarely the case.

[10] This illustration ignores likely temporal dynamics in rebels' strength, rendering Panel B a 'proxy control' scenario. This nuance will be discussed in more detail below.

[11] See, for example, Rosenbaum (1984), King, Keohane & Verba (1994), Rubin (2004), Gelman & Hill (2007), Imai, King & Stuart (2008), Angrist & Pischke (2009), Pearl (2009), VanderWeele, Mumford & Schisterman (2012), Westreich & Greenland (2013), Elwert & Winship (2014), Mayer et al. (2014), Acharya, Blackwell & Sen (2016), Pearl, Glymour & Jewell (2016), Montgomery, Nyhan & Torres (2018), Cinelli, Forney & Pearl (2020) and Groenwold, Palmer & Tilling (2021). Including a covariate in a regression model is a special case of conditioning, which also subsumes subsetting and, more generally, missingness. For simplicity, the language in this article focuses on covariate inclusion in regression models.

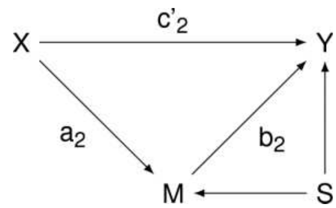

Figure 5. Basic extension

What about instances in which researchers *wish* to isolate a certain mechanism by partialling out another mechanism, or discern the magnitude of one path independent of another? Acharya, Blackwell & Sen (2016) find that this is the case for 23% of publications that condition on posttreatment variables, out of a sample of publications from three top political science journals[12] between 2010 and 2015. However, the 'isolated' direct effect c'_2 is only representative of observations for which X has no influence on M. Such observations may not exist, and the results may therefore not be reflective of reality.[13]

Yet more importantly, assuming that c'_2 was indeed the quantity of interest and there were any observations of which this was representative, causality in applied research is rarely as simple as in Figure 3. Even slight modifications to this basic setup can add significant bias to the estimate of c'_2 if this is not modelled using proper mediation analysis techniques (Imai, Keele & Tingley, 2010; Imai et al., 2011; Pearl, 2014; Hayes, 2018). For example, consider the simple extension visualized in Figure 5. Here, the variable S is added, influencing both M and Y. M is now a 'collider variable' on the path between X and S: the paths from the two variables meet in M, meaning that M blocks an effect transmission between X and S (Pearl, Glymour & Jewell, 2016). Therefore, in a simple bivariate regression of Y on X, S does not confound the total effect c_2. However, a researcher looking to isolate c'_2 by conditioning on M, without taking S into account, opens a non-causal path from X to S. The estimate of c'_2 will be biased (Rosenbaum, 1984; Imai, Keele & Yamamoto, 2010; Elwert & Winship, 2014; Acharya, Blackwell & Sen, 2016). This is a form of selection bias also known as 'collider-stratification bias.' Depending on functional form and model setup, more bias accumulates quickly (Imai, Keele & Tingley, 2010; Glynn, 2012; Pearl, 2014).

In sum, whether the aim is to estimate the total treatment effect or the direct treatment effect, simply conditioning on a posttreatment variable without considering the underlying assumptions is probably adding bias rather than reducing it. When the quantity of interest is the total treatment effect, as is the case in most empirical peace and conflict research, a variable that is purely posttreatment should just be disregarded altogether. Unfortunately, many confounders in applied research are not purely posttreatment and exhibit directional ambiguity, which is more complicated to address and will be discussed in the next section. When the interest lies with the direct treatment effect, this needs to be appropriately modelled: while, in a basic artificial setup and under strong assumptions, the coefficient estimates may still be recovered by a simple regression model, in a more realistic research scenario it is almost certainly necessary to implement a formal mediation analysis.[14] See Carter, Shaver & Wright (2019) for an application in peace and conflict studies, showing how the notorious effect of rugged terrain (X) on civil war (Y) is mediated by political marginalization (M).

Appreciating the causal sequence between variables, and the difference between the direct and indirect treatment effect, also challenges a straightforward interpretation of covariate results in multivariate regression. In an analysis that conditions on a confounder Z, the covariate's coefficient suffers from the same biases discussed above. This becomes intuitively apparent when considering Panel A in Figure 3: the total effect of Z on Y is a combination of both b_1 and a_1c_1. In an analysis that regresses Y on X and Z, the estimated coefficient of Z cannot be interpreted as the total effect of that covariate

[12] *American Political Science Review*, *American Journal of Political Science* and *World Politics*.

[13] This is the case when conditioning on M in a 'standard' regression framework, yielding the natural direct effect as coefficient estimate for c'_2 (VanderWeele, Mumford & Schisterman, 2012; Acharya, Blackwell & Sen, 2016). To relax this assumption, the controlled direct effect may be estimated using sequential *g*-estimation. See Acharya, Blackwell & Sen (2016) for a nuanced comparison of the natural and controlled direct effect.

[14] For more information on estimators and assumptions in mediation analysis, see, for example, VanderWeele (2009), Imai, Keele & Tingley (2010), Imai, Jo & Stuart (2011), Imai et al. (2011), Shpitser & VanderWeele (2011), Gerber & Green (2012); Pearl (2012), Pearl (2014) and Hayes (2018). For setups in which fully modelling the relationship between mediator and outcome is not attainable, the direct treatment effect can be estimated using either an instrumental variable approach, as shown in Aklin & Bayer (2017), or via the average controlled direct effect approach and accompanied sensitivity analysis, as discussed in VanderWeele, Mumford & Schisterman (2012) and Acharya, Blackwell & Sen (2016). For using an instrumental variable approach in an experimental setting, see Montgomery, Nyhan & Torres (2018).

on the outcome, because part of the effect of the covariate is relayed through X. Moreover, the covariate's coefficient estimate likely suffers itself from OVB, because the model specification is usually tailored towards the explanatory variable of interest and does not consider second-hand confounders. Therefore, neither the substantive effect size nor the significance of covariates' coefficient estimates can be meaningfully interpreted in a standard regression setup. The mistaken belief in such 'mutual adjustment' among covariates is also known as the 'Table 2 Fallacy' (Westreich & Greenland, 2013).[15] Nevertheless, a majority of evaluated studies on peace and conflict actively interpret their covariate results.

The total treatment effect in an imperfect world

The implication of the previous section is that all research should include a transparent discussion of the assumed relationship between treatment and covariates. Studies seeking to estimate a total treatment effect may condition on pretreatment variables, and should avoid conditioning on posttreatment variables.[16] Unfortunately, the real world of potential covariates does not only divide into pretreatment and posttreatment.

What to do when a covariate acts as both confounder and mediator? A version of this problem is exemplified by Angrist & Pischke (2009) under the name 'proxy control:' in order to account for a confounder that cannot be measured (an 'unobserved confounder;' U), the researcher has the option to use a proxy variable. That proxy is, however, only observed after the treatment. A dilemma occurs in which including the covariate in the analysis alleviates OVB because it proxies for a pretreatment confounder, but at the same time induces over-control bias because it was measured after treatment.[17]

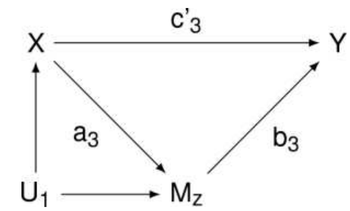

Figure 6. Proxy control

Not including it means to avoid posttreatment bias, but to risk OVB. The researcher is caught between a rock and a hard place.

One of the most eminent works in peace and conflict research, *Why Civil Resistance Works: The Strategic Logic of Nonviolent Conflict* (Chenoweth & Stephan, 2011), illustrates the challenge of accommodating a proxy control. Studying how the use of nonviolent means (X) influences resistance campaigns' likelihood of achieving their policy goals (Y), Chenoweth & Stephan (2011) argue that nonviolence attracts a larger and more diverse support base, which in turn improves the movement's likelihood of success. However, not only do the means of resistance influence a campaign's size (M), but a campaign's size also determines, both directly and indirectly, the means that it is able to adopt (Z). Chenoweth & Stephan (2011) address this issue by including peak campaign size as a covariate in their models, thereby mitigating OVB. Meanwhile, with campaign size also being a key mechanism that relays part of the effect of nonviolence on campaign success, its inclusion as a covariate is also bound to induce over-control bias.

Figure 6 visualizes an adaptation of the problem, showing the variable M_Z as an intermediate variable on the causal path between X and Y, and between the unobserved pretreatment confounder U_1 and Y.[18] In terms of the example above, X can be substituted with nonviolence and Y with campaign success, while M_Z represents the posttreatment number of participants and U_1 the pretreatment campaign size. Recalling the effect decomposition outlined earlier, it is apparent that U_1 makes an inclusion of M_Z in our model desirable lest the researcher risks OVB, while a_3 cautions against including M_Z to avoid over-control bias ($c'_3 \neq c_3$). When discussing solutions to this dilemma in his 2010 presentation, Gary King ended on the quip 'Is there hope? There's always hope; just no answers!' (King,

[15] Also see Keele, Stevenson & Elwert (2020) for a comprehensive treatment of this topic and a dissection of the identifying assumptions.

[16] While this implies that variables being pretreatment is a necessary criterion for inclusion (apart from fringe cases), it is by no means a sufficient criterion. There are many instances in which conditioning on pretreatment variables is not advisable (see e.g. M-bias). While elaborating on such and other caveats goes beyond the scope of this article, readers may turn to, for example, Elwert & Winship (2014) and Cinelli, Forney & Pearl (2020) for an overview of various variable constellations.

[17] This is in addition to endogenous selection bias by opening a noncausal path through the unobserved confounder, and residual omitted variable bias due to being a mere proxy measure of the original confounder. See Elwert & Winship (2014) for a detailed review of this and other scenarios.

[18] Instead of U_1, the same problem can be represented as X and M_Z exhibiting a simultaneous relationship, with path a_3 going in both directions (Pearl, 2009; Bellemare, Masaki & Pepinsky, 2017).

2010). While there are still no 'answers,' in the following I highlight possible avenues to confront the issue, and limitations thereof, for empirical research on peace and conflict.

Sensitivity analysis. A canonical way of addressing the dilemma of a proxy control is to conduct a sensitivity analysis, as recently recommended by Groenwold, Palmer & Tilling (2021) and dating back to Rosenbaum (1984). The basic logic behind a sensitivity analysis is to examine the results' robustness under various assumptions about the setup and strengths of association between variables. While this is not a 'fix,' it enables the researcher to better understand the assumptions under which their findings hold and to exercise transparency about the limitations of their study. Sensitivity analyses may be conducted via bias expressions (analytically) or via simulation (computationally).

Computational sensitivity analyses are easily achievable given the common provision of powerful statistical software.[19] Groenwold, Palmer & Tilling (2021) suggest several *R* packages on structural equation modelling that include useful functionality for simulating data, including *DAGitty* (Textor et al., 2016) and *lavaan* (Rosseel, 2012). *DAGitty* also comes with a browser-based environment to ease the introduction to, and interaction with, graphical representations of model specifications (dyadic acyclic graphs; DAGs) at *www.dagitty.net*. A recent tool worth highlighting is *sensemakr* by Cinelli & Hazlett (2020), which was specifically designed for sensitivity analysis in applied research. It offers comprehensive and easy-to-use functionality in both *R* and STATA for evaluating OVB and visualizing bounds via contour plots (Cinelli, Ferwerda & Hazlett, 2020).

Package selection depends on the researcher's needs and preferences, but the procedure when estimating the total treatment effect is similar across tools. In a first step, the researcher estimates their regression model using only pretreatment covariates and excluding potentially offending covariates – that is, excluding all covariates that may be directly or indirectly influenced by the explanatory variable of interest. This risks OVB but avoids over-control and collider-stratification bias. In a second step, the researcher conducts an analysis of their treatment effect's sensitivity to OVB using one of the aforementioned tools, and reports the findings (e.g. through a contour plot) together with their regression results. The material cited above guides readers through the implementation and interpretation of the sensitivity analysis. In this context I specifically highlight *sensemakr* by Cinelli & Hazlett (2020) for providing a very accessible guide. Among a small but growing number of studies on peace and conflict that use *sensemakr* to probe their results' sensitivity to OVB, see, for example, Koos & Lindsey (2022) and Pinckney, Butcher & Braithwaite (2022) for applications.

As a form of manual sensitivity analysis, some empirical research attempts to bound the total treatment effect by estimating two regressions, one that includes M_Z and one that does not, hoping that the true parameter may lie somewhere in-between. Unfortunately, this is not generally the case (King, 2010; Groenwold, Palmer & Tilling, 2021). As methodological literature lacks a systematic and accessible treatment of this approach, I provide an extended discussion and proofs in the Online appendix.[20] Such a bounding exercise implicitly assumes that the total treatment effect suffers from downward bias in one regression and upward bias in the other. In other words, it would require a scenario in which (not) conditioning on M_Z leads to strictly positive (negative) bias, or vice versa. Determining the direction of bias can also be useful beyond bounding, especially in the context of a null-hypothesis test, as a simple way of investigating whether the estimated effect likely over-estimates or under-estimates the true relationship. However, while recognizing the direction of bias solely based on assumed effect signs (positive/negative) can be trivial in an artificial setup such as Figure 3, it can quickly become impossible in a more realistic scenario.

Based on general bias expressions in Groenwold, Palmer & Tilling (2021), I derive conditions under which bias direction is identifiable solely based on effect signs, and in the context of a more realistic setting that combines all the challenges discussed so far: a mediator variable that also serves as proxy control, paired with S that makes M_Z a collider. This allows identifying scenarios in which the total treatment effect can be bounded irrespective of relative effect sizes, pitting OVB against posttreatment bias and collider-stratification bias. These

[19] To increase accessibility, I solely focus on computational approaches that rely on ready packages. Also, due to the focus on the total treatment effect, I do not include discussions of literature specific to sensitivity analyses for collider-stratification bias in the estimation of direct effects (see e.g. Whitcomb et al. (2009) for an application).

[20] For more information on analytical solutions and manually calculated bias expressions, see also Rosenbaum (1984), VanderWeele & Arah (2011), VanderWeele, Mumford & Schisterman (2012) and Blackwell (2014).

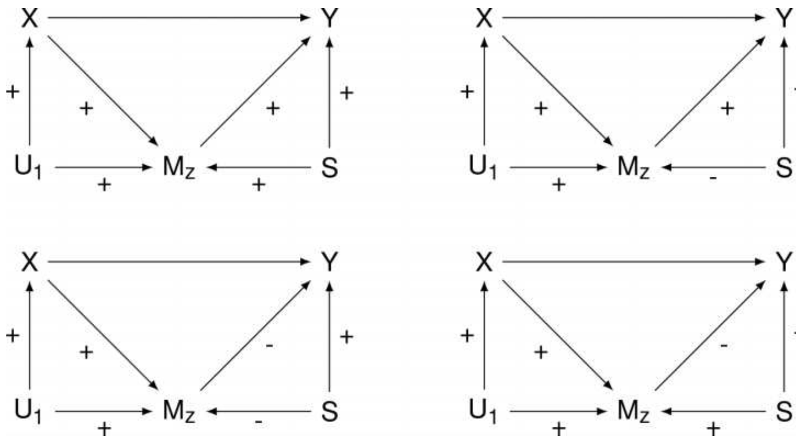

Figure 7. Combinations of effect directions that allow bounding

few scenarios in which the true total effect lies strictly in-between the estimates from the two regressions are visualized in Figure 7.[21]

This is useful for two reasons. First, Figure 7 allows a researcher to map their assumed effect directions in their own theoretical model setup onto these graphs. If they match, this may enable the researcher to identify the upper and lower limit of their total treatment effect through a bounding exercise. Second, Figure 7 is useful due to all the combinations and settings it does not display. In other words, it serves as a reminder that conducting two regressions and claiming they bound the true effect without carefully considering the underlying assumptions is, in most instances, wrong. Even slight changes to these setups can result in different dynamics that render bias direction dependent on relative effect sizes (Elwert & Winship, 2014; Cinelli, Forney & Pearl, 2020; Groenwold, Palmer & Tilling, 2021). Therefore, when arguing for a certain bias direction, researchers must exercise caution and be transparent regarding their assumptions. Finally, it is important to recall that even when the total treatment effect is successfully bounded, its possible realizations are not uniformly distributed between the two limits and may cluster close to either boundary.[22]

The Online appendix section 2 provides further analytical results that can help to probe empirical findings in a more nuanced way (see Online Equations 2.2a and 2.5). Based on these expressions, researchers can explicitly state their assumed data generating process to examine the direction of bias in their projects, whether their regression coefficient may be a conservative or inflated estimate, as well as whether their total treatment effect can be bounded. This discussion of the opportunities for and limitations of inferring the direction of bias should not discourage researchers from reporting different model specifications. While special attention should be given to a careful interpretation and prudent inference, it is always good practice to exercise transparency over the results' model dependence. To this end, Young & Holsteen (2017) developed a systematic framework for estimating and reporting model uncertainty and robustness in the context of different specifications.

Lagging covariates. The use of panel data in peace and conflict studies is ubiquitous, presenting unique challenges and opportunities. Research on social science methodology made important advances on the practice of lagging as a way to ameliorate reverse causality between treatment and outcome as a source of endogeneity.[23] Lagging a covariate as a way of addressing the proxy control problem has not received systematic

[21] Online appendix section 2 walks through the derivations of the analytical results. For consistency of nomenclature across literature, the Online appendix denotes S as U_2.

[22] Visual examples for this behaviour can be found in the Online appendix section 2.

[23] A temporally lagged variable is a variable measured at an earlier time point, typically denoted as t-1 (or t-2, t-3, etc. depending on the order of the lag). For example, if the data are observed at the yearly level, a lagged variable of first order takes the values of the year before. In brief, the popularity of lagging a treatment as an all-in-one solution to address reverse causality is not supported by recent findings; see, for example, Bellemare, Masaki & Pepinsky (2017) and Blackwell & Glynn (2018. In particular, in the face of interrelated time-dependent processes, temporal order of measurement does not usually translate into causal order.

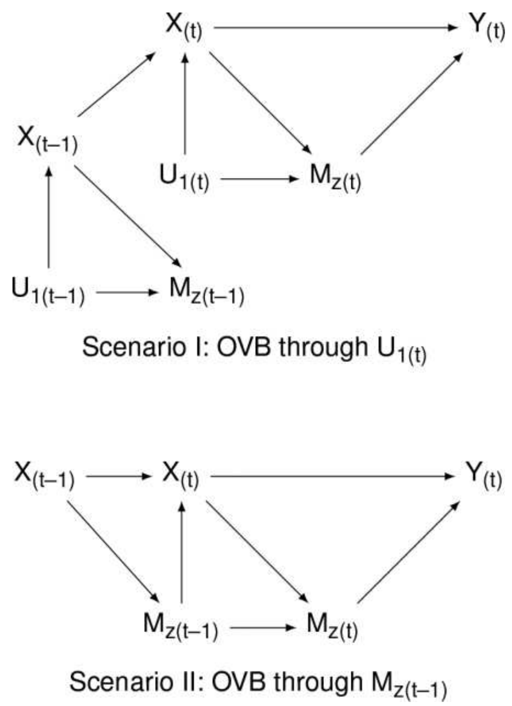

Figure 8. Proxy control (panel)

attention in methodological research, and the number and the heterogeneity of possible setups warrants a separate study. However, due to the ubiquity of panel data in peace and conflict research, I offer an abridged discussion of two exemplary data generating processes in a panel structure for which solutions can be inferred from related literature.

Figure 8 visualizes two scenarios that help to exemplify the potential use and limitations of lagging to address the proxy control problem. Scenario I offers a simple extension of Figure 6, adding temporal dependence to the treatment X. In this particular case, the total treatment effect can be estimated by using past values of X and not conditioning on M_Z. This is because in a regression of $Y_{(t)}$ on $X_{(t-1)}$, the latter is exogenous to $U_{1(t)}$. However, as with the bias expressions described above, the main lesson lies in what Scenario I does *not* show: there are no connections between $M_{Z(t-1)}$ and $M_{Z(t)}$, or between $U_{1(t-1)}$ and $U_{1(t)}$. In other words, the scenario assumes no temporal dynamics for U_1 and M_Z. Depending on the applied context, this is a very strong assumption that, if violated, results in bias. In particular, lagging X when there is time dependence in U_1 can make matters substantially worse compared to no lagging (Bellemare, Masaki & Pepinsky, 2017).

In Scenario II, past values of M_Z influence both current values of M_Z and current values of X. This scenario exemplifies when the proxy control problem is caused by temporal dynamics in the covariate, and $M_{Z(t-1)}$ takes the role of what was previously $U_{1(t)}$. Therefore, this is an example for a case in which the panel structure makes the relationship between X and M_Z – which seems simultaneous in a time invariant (one shot) setting – unfurl over time. This renders the endogeneity acyclic and fully observed, and the proxy control problem becomes more easily approachable. The total treatment effect can be recovered by simply regressing $Y_{(t)}$ on $X_{(t)}$ and conditioning on $M_{Z(t-1)}$ (cf. Pearl 2009). An applied example of this solution in peace and conflict research is provided by Smidt (2020). Studying the effect of election education ($X_{(t-1)}$) on electoral violence ($Y_{(t)}$), Smidt (2020) identifies a location's baseline violence as an important confounding factor that influences both the allocation of election education and future levels of violence. Instead of measuring baseline violence during the same time as election education events take place ($M_{Z(t)}$), which would risk that baseline violence is influenced by election education, Smidt (2020) recognizes the potential for posttreatment bias and measures baseline violence in the months before any education events take place ($M_{Z(t-1)}$). Awareness and transparency go a long way in improving observational causal inference, and this example is illustrative of how easy it can be to mitigate posttreatment bias. As always, the viability of this solution is dependent on the underlying model setup that the researcher assumes. Simply lagging the treatment or a covariate without considering why this may mitigate bias in a project-specific theoretical and empirical context is not tenable.

Total effect decomposition (TED). Finally, a potential solution to the proxy control problem is offered by the TED approach developed by Aklin & Bayer (2017). It is an intuitive way to recover unbiased effect estimates for the case of a binary treatment (dummy) variable ($X = \{0, 1\}$), assuming linearity and treatment effect homogeneity. While I focus on the total treatment effect, enabling me to narrow the discussion down to the first three steps of the approach, TED is designed to also recover the direct and indirect effects in subsequent steps. Similar to the regression preceding a sensitivity analysis, the offending variable M_Z is left out in this approach. It uses only pretreatment covariates (Z) that must be unrelated to M_Z to predict counterfactuals of the outcome. If these conditions are met, the total treatment effect can be estimated in a three-step procedure:

1. The outcome Y is regressed on all covariates that are pretreatment (all confounders Z; these must be unrelated to any posttreatment variable M_Z), but only for the subset of untreated observations (X = 0). Also written as regressing: $Y_{X=0} = \beta_0 + \beta_1 Z_{X=0} + \epsilon$
2. Using the estimates from the previous step, predicted values are calculated for all treated observations (X = 1). Also written as predicting: $\hat{Y}_{X=1} = \hat{\beta}_0 + \hat{\beta}_1 Z_{X=1}$
3. The total effect estimate c is the mean difference between the predicted values of the previous step ($\hat{Y}_{X=1}$) and the observed outcome values among the treated $Y_{X=1}$. Also written as calculating: $c = E[\hat{Y}_{X=1} - Y_{X=1}]$. Uncertainty intervals can be estimated via bootstrapping.

To summarize, addressing the challenge of a proxy control requires nuance. When having to choose between OVB and posttreatment bias, there is no all-in-one solution: depending on the theoretical framework and empirical setup, different conditions and assumptions apply. In all instances, however, exercising transparency and prudence in the interpretation of results is key. To this end, sensitivity analyses are instrumental to the quantification and communication of uncertainty.

Conclusion and implications for applied research

The peace and conflict research programme has continuously improved its application of quantitative methods. Works such as Clarke (2009) and Schrodt (2013), and seminal contributions by Gary King, Kosuke Imai, and others, substantially contributed to scholars' understanding and awareness. Prominently, this included the slow abandonment of garbage can models (cf. Achen, 2005). However, while the notion of 'more covariates is always better' became rightfully outdated, less attention was given to the *suitability* of those covariates that remained and make up today's models in peace and conflict research. Many studies justify the inclusion of covariates merely based on their relationship with the outcome, or worse, by declaring them the 'usual set of controls' – the inclusion of which is, oftentimes, either pointless or detrimental for minimizing bias in estimation.

There is much to consider in the context of research design and model specification, a discussion of which goes beyond the scope of any individual article.[24] I focus on one issue that can help to significantly improve estimation practice in peace and conflict studies. Manuscripts must give more emphasis to the discussion of model specification, explaining how each covariate relates not only to the outcome, but also to the explanatory variable of interest (the treatment). In doing so, the direction of these relationships requires attention: does the covariate influence the treatment, or does the treatment influence the covariate?

The direction of the effect between the treatment and each covariate matters. Conditioning on a posttreatment variable, that is, a covariate preceded by the treatment, biases the total treatment effect estimate. Meanwhile, even if the aim is to isolate an individual mechanism (direct treatment effect), due care has to be given to the modelling strategy and underlying assumptions to mitigate bias. The solution for avoiding posttreatment bias is easy: not to include posttreatment variables in one's model specification. However, there are variables that may be partly influenced by the treatment, but at the same time also proxy for an exogenous confounder. Not conditioning on them means to avoid posttreatment bias, but to risk OVB – conditioning on them means to account for OVB, but to accept posttreatment bias. These 'proxy controls' require a careful assessment and transparent discussion. A prudent way to address such a dilemma is to conduct a computational sensitivity analysis that quantifies the extent to which research findings are sensitive to OVB. Other potential avenues, depending on the research setup and the underlying assumptions, are to conduct an analytical sensitivity analysis, lag covariates to temporally precede the treatment variable, or follow the TED approach.

[24] Article-length reviews of important core topics can be found in the *Conflict Management and Peace Science* 2005 special issue (Kadera & Mitchell, 2005), Clarke (2009), Schrodt (2013), Elwert & Winship (2014) and Cinelli, Forney & Pearl (2020). Among the many excellent textbooks that provide a more comprehensive background reading, I would emphasize Cunningham (2021), Huntington-Klein (2021) and Llaudet & Imai (2022) as being particularly accessible teaching resources. The same applies to Angrist & Pischke (2009), Imai (2018) and McElreath (2020) at a more advanced level.

Abridged list of recommendations[25] for how to avoid posttreatment bias when estimating a total treatment effect

- For each variable that is to be included as a covariate ('control') in the model, ask: does it influence the explanatory variable of interest (treatment), or the other way around?
  - Treatment ← variable: likely a covariate, may be included (pp. 5-6).
  - Treatment → variable: likely a mediator, should not be included (pp. 6-8).
  - Treatment ↔ variable: likely a proxy control (pp. 8-9). Approaches and limitations:
    1. Computational sensitivity analysis (p. 9). Not including the proxy control and assessing the treatment effect's sensitivity to OVB serves most applications. This recommendation, however, assumes precision and reliability in conducting the sensitivity analysis. To this end, Cinelli & Hazlett (2020) provide accessible guides.[26]
    2. Bounding (pp. 9-10). Identifying a range in which the treatment effect lies by running two analyses, one that includes and one that excludes the proxy control, only works for specific setups and requires a discussion of all assumed effect directions.
    3. Lagging (pp. 10-11). Just as the other approaches, this is no panacea. In particular, in the case of temporal dependencies, it requires a careful discussion of the data generating process.
    4. TED (pp. 11-12). This approach may help depending on the nature of the treatment and the overall data generating process.
- Estimating a direct treatment effect usually requires a formal mediation analysis (p. 7).
- Covariate coefficients usually do not warrant interpretation (pp. 7-8).

I show that a majority of quantitative hypothesis tests in recent publications in the field of peace and conflict research may be biased due to a lack of consideration for the issue of posttreatment variables. Just as reverse causality between the treatment and the outcome is an important point of consideration for authors, reviewers and editors, causal direction between the treatment and covariates should be as well. Moreover, a majority of evaluated studies actively interpret their covariates' effect estimates or significance levels, suggesting a widespread lack of awareness for the role of model specification. It is this awareness that this study seeks to raise, in an effort to facilitate transparent discussions on variable selection in the peace and conflict research programme. Questions surrounding research design and model specification are never easy to answer, but they are important to ask.

Replication data

The dataset, codebook, and script for the article review, along with the Online appendix, are available at https://www.prio.org/jpr/datasets/ and on https://www.chrisdworschak.com/research.html. Data management was conducted using R.

Acknowledgements

I thank the *Journal of Peace Research* editors and two anonymous reviewers for their helpful feedback, as well as Patrick Bayer, Charles Butcher, Daina Chiba, Amélie Godefroidt, Thea Johansson, Moritz Marbach, Phillip Nelson, Clara Neupert-Wentz, Espen Geelmuyden Rød, Christoph V Steinert and Felix B Weber for many useful comments and conversations that helped shape this article. Moreover, I am thankful for great suggestions by the Norwegian University of Science and Technology peace research group, participants at the University of Essex' Michael Nicholson Centre for Conflict and Cooperation workshop and the Interdisciplinary Peace and Conflict Research Network workshop, as well as by the peace researchers at the 2021 annual conferences of the Conflict Research Society and of the German Association for Peace and Conflict Studies. Special thanks to Christoffer Andersen, Lasse Holtar, Conor Kelly, Andreas Lillebråten and Niclas Weischner for their excellent research assistance. All mistakes are my own.

[25] I thank the editor and an anonymous reviewer for their suggestion to include this "list of recommendations" to further improve accessibility. It succinctly illustrates the importance of distinguishing between covariates, mediators, and proxy controls – a distinction that still finds only limited appreciation in the peace and conflict research programme. Naturally, condensing this article's discussion into such an abridged checklist comes at the expense of nuance. Therefore, this list should be considered a map rather than a tl;dr summary. When implementing any of the approaches, please use the page numbers (in brackets) for details, caveats, and reference to further literature.

[26] For example, see their vignette at https://cran.r-project.org/web/packages/sensemakr/vignettes/sensemakr.html.

Funding

The author received no independent financial support for the research, authorship, and/or publication of this article.

ORCID iD

Christoph Dworschak https://orcid.org/0000-0003-0196-9545

References

Acharya, Avidit; Matthew Blackwell & Maya Sen (2016) Explaining causal findings without bias: Detecting and assessing direct effects. *American Political Science Review* 110(3): 512–529.

Achen, Christopher H (2005) Let's put garbage-can regressions and garbage-can probits where they belong. *Conflict Management and Peace Science* 22(4): 327–339.

Aklin, Michaeël & Patrick Bayer (2017) How can we estimate the effectiveness of institutions? Solving the post-treatment versus omitted variable bias dilemma. *SemanticScholar*. Working paper. (https://www.semanticscholar.org/paper/How-Can-We-Estimate-the-Effectiveness-of-Solving-Aklin-Bayer/b764596a4dae23b719df2816d84e08f618f4e641).

Angrist, Joshua D & Joörn-Steffen Pischke (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Bellemare, Marc F; Takaaki Masaki & Thomas B Pepinsky (2017) Lagged explanatory variables and the estimation of causal effect. *Journal of Politics* 79(3): 949–963.

Blackwell, Matthew (2014) A selection bias approach to sensitivity analysis for causal effects. *Political Analysis* 22(2): 169–182.

Blackwell, Matthew & Adam N Glynn (2018) How to make causal inferences with time-series cross-sectional data under selection on observables. *American Political Science Review* 112(4): 1067–1082.

Carter, David B; Andrew C Shaver & Austin L Wright (2019) Places to hide: Terrain, ethnicity, and civil conflict. *Journal of Politics* 81(4): 1446–1465.

Chenoweth, Erica & Maria J Stephan (2011) *Why Civil Resistance Works: The Strategic Logic of Nonviolent Conflict*. New York: Columbia University Press.

Cinelli, Carlos; Jeremy Ferwerda & Chad Hazlett (2020) sensemakr: Sensitivity analysis tools for OLS in R and Stata. *SSRN* Working paper. (https://papers.ssrn.com/sol3/papers.cfm?Abstract_id=3588978).

Cinelli, Carlos; Andrew Forney & Judea Pearl (2020) A crash course in good and bad controls. *SSRN* Working paper. DOI: https://dx.doi.org/10.2139/ssrn.3689437.

Cinelli, Carlos & Chad Hazlett (2020) Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society* 82(1): 39–67.

Clarke, Kevin A (2005) The phantom menace: Omitted variable bias in econometric research. *Conflict Management and Peace Science* 22(4): 341–352.

Clarke, Kevin A (2009) Return of the phantom menace: Omitted variable bias in political research. *Conflict Management and Peace Science* 26(1): 46–66.

Cunningham, Scott (2021) *Causal Inference: The Mixtape*. London: Yale University Press.

Elwert, Felix & Christopher Winship (2014) Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology* 40(1): 31–53.

Gelman, Andrew & Jennifer Hill (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models (Analytical Methods for Social Research)*. Cambridge: Cambridge University Press.

Gerber, Alan S & Donald P Green (2012) *Field Experiments: Design, Analysis, and Interpretation*. New York: WW Norton & Company.

Glynn, Adam N (2012) The product and difference fallacies for indirect effects. *American Journal of Political Science* 56(1): 257–269.

Groenwold, Rolf HH; Tom M Palmer & Kate Tilling (2021) To adjust or not to adjust? When a "confounder" is only measured after exposure. *Epidemiology* 32(2): 194–201.

Hayes, Andrew F (2018) *Introduction to Mediation, Moderation, and Conditional Process Analysis, Second Edition: A Regression-Based Approach (Methodology in the Social Sciences)*. New York: Guilford.

Hernaán, Miguel A (2018) The c-word: Scientific euphemisms do not improve causal inference from observational data. *American Journal of Public Health* 108(5): 616–619.

Huntington-Klein, Nick (2021) *The Effect: An Introduction to Research Design and Causality*. Abingdon: Routledge.

Imai, Kosuke (2018) *Quantitative Social Science: An Introduction*. Princeton, NJ: Princeton University Press.

Imai, Kosuke; Booil Jo & Elizabeth A Stuart (2011) Commentary: Using potential outcomes to understand causal mediation analysis. *Multivariate Behavioral Research* 46(5): 861–873.

Imai, Kosuke; Luke Keele & Dustin Tingley (2010) A general approach to causal mediation analysis. *Psychological Methods* 15(4): 309–334.

Imai, Kosuke; Luke Keele, Dustin Tingley & Teppei Yamamoto (2011) Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review* 105(4): 765–789.

Imai, Kosuke; Luke Keele & Teppei Yamamoto (2010) Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science* 25(1): 51–71.

Imai, Kosuke; Gary King & Elizabeth A Stuart (2008) Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society* 171(2): 481–502.

Kadera, Kelly M & Sara Mclaughlin Mitchell (2005) Manna from heaven or forbidden fruit? The (ab) use of control variables in research on international conflict. *Conflict Management and Peace Science* 22(4): 273–275.
Keele, Luke; Randolph T Stevenson & Felix Elwert (2020) The causal interpretation of estimated associations in regression models. *Political Science Research and Methods* 8(1): 1–13.
King, Gary (2010) A hard unsolved problem? Post-treatment bias in big social science questions. *Harvard University. Presented at the "Hard Problem in Social Science" Symposium, Harvard University, 4/10/2010*. (https://gking.harvard.edu/presentations/hard-unsolved-problem-post-treatment-bias-big-social-science-questions).
King, Gary; Robert O Keohane & Sidney Verba (1994) *Designing Social Inquiry*. Princeton, NJ: Princeton University Press.
Koos, Carlo & Summer Lindsey (2022) Wartime sexual violence, social stigmatization and humanitarian aid: Survey evidence from eastern Democratic Republic of Congo. *Journal of Conflict Resolution* 66(6): 1037–1065.
Llaudet, Elena & Kosuke Imai (2022) *Data Analysis for Social Science*. Princeton, NJ: Princeton University Press.
Mayer, Axel; Felix Thoemmes, Norman Rose, Rolf Steyer & Stephen G West (2014) Theory and analysis of total, direct, and indirect causal effects. *Multivariate Behavioral Research* 49(5): 425–442.
McElreath, Richard (2020) *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*. London: Chapman and Hall/CRC.
Montgomery, Jacob M; Brendan Nyhan & Michelle Torres (2018) How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science* 62(3): 760–775.
Pearl, Judea (2009) *Causality: Models, Reasoning, and Inference*. 2nd edition. Cambridge: Cambridge University Press.
Pearl, Judea (2012) The causal mediation formula – a guide to the assessment of pathways and mechanisms. *Prevention Science* 13(4): 426–436.
Pearl, Judea (2014) Interpretation and identification of causal mediation. *Psychological Methods* 19(4): 459–481.
Pearl, Judea; Madelyn Glymour & Nicholas P Jewell (2016) *Causal Inference in Statistics: A Primer*. Chichester: Wiley.
Pinckney, Jonathan; Charles Butcher & Jessica Maves Braithwaite (2022) Organizations, resistance, and democracy: How civil society organizations impact democratization. *International Studies Quarterly* 66(1): sqab094.
Rosenbaum, Paul R (1984) The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society* 147(5): 656–666.
Rosseel, Yves (2012) lavaan: An R package for structural equation modeling. *Journal of Statistical Software* 48(1): 1–36.
Rubin, Donald B (2004) Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics* 31(2): 161–170.
Schrodt, Philip A (2013) Seven deadly sins of contemporary quantitative political analysis. *Journal of Peace Research* 51(2): 287–300.
Shpitser, Ilya & Tyler J VanderWeele (2011) A complete graphical criterion for the adjustment formula in mediation analysis. *International Journal of Biostatistics* 7(1): Article 16.
Smidt, Hannah (2020) Mitigating election violence locally: UN peacekeepers' election-education campaigns in Côte d'Ivoire. *Journal of Peace Research* 57(1): 199–216.
Textor, Johannes; Benito van der Zander, Mark S Gilthorpe, Maciej Liskiewicz & &George T H Ellison (2016) Robust causal inference using directed acyclic graphs: The R package 'dagitty'. *International Journal of Epidemiology* 45(6): 1887–1894.
VanderWeele, Tyler J (2009) Marginal structural models for the estimation of direct and indirect effects. *Epidemiology* 20(1): 18–26.
VanderWeele, Tyler J & Onyebuchi A Arah (2011) Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology* 22(1): 42–52.
VanderWeele, Tyler J; Sunni L Mumford & Enrique F Schisterman (2012) Conditioning on intermediates in perinatal epidemiology. *Epidemiology* 23(1): 1–9.
Westreich, Daniel & Sander Greenland (2013) The table 2 fallacy: Presenting and interpreting confounder and modifier coefficients. *American Journal of Epidemiology* 177(4): 292–298.
Whitcomb, Brian W; Enrique F Schisterman, Neil J Perkins & Robert W Platt (2009) Quantification of collider-stratification bias and the birthweight paradox. *Paediatric and Perinatal Epidemiology* 23(5): 394–402.
Wood, Reed M; Jacob D Kathman & Stephen E Gent (2012) Armed intervention and civilian victimization in intrastate conflicts. *Journal of Peace Research* 49(5): 647–660.
Young, Cristobal & Katherine Holsteen (2017) Model uncertainty and robustness: A computational framework for multimodel analysis. *Sociological Methods & Research* 46(1): 3–40.

CHRISTOPH DWORSCHAK, PhD in Government (University of Essex, 2020); Assistant Professor, University of York (2022–); Fellow, Michael Nicholson Centre for Conflict and Cooperation, University of Essex (2020–); Postdoc, Norwegian University of Science and Technology (2021–2022) and University of the German Armed Forces Munich (2020–2021); Visiting Researcher, University of Mannheim (2020), Princeton University (2019) and Peace Research Institute Oslo (2019); conducts research on repression and dissent, peacekeeping and counterinsurgency, and civil–military relations.