From "Is it unconfounded?" to "How much confounding would it take?": Applying the sensitivity-based approach to assess causes of support for peace in Colombia

Chad Hazlett*, Francesca Parente†

Abstract

Attention to the credibility of causal claims has increased tremendously in recent years. When relying on observational data, debate often centers on whether investigators have ruled out *any* bias due to confounding. However, the relevant scientific question is generally not whether bias is precisely zero, but whether it is problematic enough to alter one's research conclusion. We argue that sensitivity analyses would improve research practice by showing how results would change under plausible degrees of confounding, or equivalently, by revealing what one must argue about the strength of confounding to sustain a research conclusion. This would improve scrutiny of studies in which non-zero bias is expected, and of those where authors argue for zero bias but results may be fragile to confounding too weak to be ruled out. We illustrate this using off-the-shelf sensitivity tools to examine two potential influences on support for the FARC peace agreement in Colombia.

*Associate Professor of Statistics and Political Science, University of California Los Angeles

†Assistant Professor of Political Science, Christopher Newport University.

1 Introduction

Many important research questions cannot be answered by randomized experiments, whether because researchers cannot ethically or practically randomize the treatments involved, or because they would like to study the effect of real world events that have already occurred. Observational alternatives to randomized experiments involve demanding and typically untestable assumptions. The most commonly employed strategies rely on an assumption of "no unobserved confounding" and then adjust for observables using estimators such as stratification, regression, matching, or various weighting approaches. Other strategies that seek to limit confounding biases, such as difference-in-difference, regression discontinuity, instrumental variables, and quasi- or natural-experiments, offer opportunities to improve inference, but are not always available and require their own demanding, largely untestable assumptions.

Fortunately, recent increases in attention to the credibility of observational findings have heightened many researchers' awareness of the risks posed by confounding. In studies where investigators adjust for observables, debate typically centers on whether *any* unobserved confounding may remain. Such "all or nothing" arguments about the threat of confounding have been a useful antidote to practices of turning a blind eye to confounding. However, we argue that this approach is suboptimal and sometimes problematic as a means of judging the risk that a result is driven by confounding. First, when authors argue for a total absence of unobserved confounding, readers and critics can and should be able to disagree, e.g. by proposing unobserved confounders that have not been ruled out. A means of debating what damage would be done by such confounders is thus required. Second and more fundamentally, the scientifically relevant concern is not usually "is there exactly zero confounding bias?", but rather "might confounding have altered the substantive research conclusion?" Yet, the arguments deployed to claim that an observational study is confounding-free are typically silent on the question of how fragile the result is in the face of possible errors in that judgment.

Consequently, even a study that is arguably confounding-free (e.g. on the basis of a convincing natural experiment) could be sensitive to very weak confounders—perhaps too small to be definitively ruled out once confronted with this information. On the other hand, studies in which the authors recognize confounding may exist could in fact require powerful confounding to alter their conclusions—perhaps so powerful that such confounding can be deemed implausible.

Many studies will fall in between these extremes. In all cases, reporting how much confounding it would take to overturn a given result is transparency-enhancing. While conventional statistics like the t-statistic, coefficient, or p-value cannot correctly answer this question, easy-to-compute and interpret "sensitivity statistics" can (Cinelli and Hazlett, 2020).

We employ off-the-shelf sensitivity analysis tools to demonstrate the transition from "all or nothing" claims about confounding to practices that reveal and communicate how much confounding it would take to change the research conclusion. For concreteness, we focus on an application examining two proposed influences on civilian voting behavior in the 2016 peace deal with the FARC in Colombia. This question has attracted great scholarly interest, with at least 14 observational papers since 2017 (see Appendix A for an expanded review). Scholars seeking to explain this real-world event have coalesced around two potential influences on voting behavior: (1) *exposure to violence* and (2) *political affiliation* with the deal's champion, President Santos. Their findings have been remarkably consistent in supporting both the exposure to violence (e.g. Tellez, 2019) and political affiliation hypotheses (e.g. Krause, 2017). Yet, none of these papers has been positioned to argue that their result is due to a causal effect rather than confounding.[1] We find that even very weak confounders could explain away the relationship between exposure to violence and FARC support. However, the effect of political affiliation requires very large confounding to overturn. Such analyses do not settle the issue of causality, but instead transparently communicate what assumptions about the strength of confounding would need to hold to achieve a given result, enabling more reasoned and productive debate than the "all or nothing" approach alone.

2 The sensitivity-based approach

We use the term sensitivity analysis specifically to refer to statistical analyses that characterize how a result changes under postulated unobserved confounding. Such tools have a long history dating back to at least Cornfield et al. (1959). Yet they remain largely unutilized: In the top three general interest political science journals in 2019, only 7 out of 119 (5.9%) papers whose conclusions may be altered by unobserved confounding included formal analyses describing the strength of confounding required to change the conclusion (see Appendix B).

[1] Experimental alternatives that seek to test particular claims are of course possible; see, e.g. Matanock and Garbiras-Díaz, 2018. However, our interest here lies in practices of making causal claims from observed, non-experimental data.

Fortunately, sensitivity statistics can be computed that correctly characterize how an estimate would change under varying degrees of postulated confounding, or equivalently, the strength of confounding required to achieve a given change in an estimate. Any sensitivity analysis approach can be useful if it (i) correctly and clearly illuminates the strength of confounding required to alter the research conclusion, and (ii) aids researchers, as cogently as possible, in understanding that strength of confounding or in debating whether such confounding plausibly exists. While similar analyses could and should be conducted for other estimation approaches (e.g. matching and weighting estimators), we focus here on the sensitivity of linear regression models because (i) linear regression is ubiquitous in political science (including in the works we consider in this application); and (ii) because the linearity of these approaches proves especially convenient for analyzing sensitivity. Of the sensitivity toolkits developed for linear regression, we employ Cinelli and Hazlett (2020), which elaborates on the concept of omitted variables bias that will be familiar to many readers. This approach also conveniently produces summary statistics to characterize sensitivity, and offers rigorous tools for comparing hypothetical confounders to observed benchmarks. We discuss the broader landscape of sensitivity approaches and how they compare in Appendix B.

While space prevents a full review of sensitivity analysis, we provide here the minimal technical material required to discuss and responsibly employ these tools. We refer readers interested in the technical details to Cinelli and Hazlett (2020) as well as to Appendix C, where we discuss finer points and commonly raised questions and concerns regarding these tools.

The key to understanding this approach begins with considering "the regression you wish you ran". Suppose that the regression of an outcome (Y) on the treatment (D) would be biased by many possible confounders, but those confounders would be controlled for if we included covariates X *and* Z (both containing one or more variables). In other words, controlling for X and Z together would render the D-Y relationship unconfounded. It would be desirable to then run the regression

$$Y = \hat{\tau} D + \boldsymbol{X}\hat{\beta} + \hat{\gamma} Z + \hat{\epsilon}_{\text{full}} \tag{1}$$

to approximate the conditioning on X and Z and provide a linear summary of the remaining relationship between D and Y. Unfortunately, while the variables in X are observed, those in

Z are unobserved. The regression actually run is instead the "restricted" regression,

$$Y = \hat{\tau}_{\text{res}} D + \boldsymbol{X}\hat{\beta}_{\text{res}} + \hat{\epsilon}_{\text{res}}. \tag{2}$$

The central question is how the observed estimate ($\hat{\tau}_{\text{res}}$) differs from the desired one, $\hat{\tau}$. We thus define $\widehat{\text{bias}} := \hat{\tau}_{\text{res}} - \hat{\tau}$, the difference between the estimate *actually* obtained and what would have been obtained in the same sample had the missing covariate(s) Z been included. As shown in Cinelli and Hazlett (2020), the bias due to omission of Z can be written as

$$|\widehat{\text{bias}}| = \text{se}(\hat{\tau}_{\text{res}}) \sqrt{\frac{R^2_{Y\sim Z|\boldsymbol{X},D}\, R^2_{D\sim Z|\boldsymbol{X}}}{1 - R^2_{D\sim Z|\boldsymbol{X}}}(\text{df})}, \tag{3}$$

The two sensitivity parameters here are partial R^2 values, $R^2_{Y\sim Z|\boldsymbol{X},D}$ and $R^2_{D\sim Z|\boldsymbol{X}}$. The first ($R^2_{Y\sim Z|\boldsymbol{X},D}$) describes how the fraction of Y *not already (linearly) explained by* X *and* D can be (linearly) explained by Z. The second, $R^2_{D\sim Z|\boldsymbol{X}}$ is similarly the fraction of the treatment status explained by confounding, after accounting for the observed covariates. A similar expression is available describing the adjusted standard error (see Cinelli and Hazlett, 2020) in terms of the same parameters, $R^2_{Y\sim Z|\boldsymbol{X},D}$ and $R^2_{D\sim Z|\boldsymbol{X}}$. A fundamental fact conveyed by these formulas is that these two parameters jointly characterize the only properties we need to know about an omitted confounder in order to determine how the point estimate, standard error, t-statistic, or p-value would be changed by including that variable.

To communicate the fragility of a result in the face of unobserved confounding we employ summary statistics described in Cinelli and Hazlett (2020), as detailed in Table 1. Additionally, "benchmarking" tools provide one useful way to interpret sensitivity analyses and argue for bound on confounding, by comparing the strength of confounding require to change the result to the explanatory power of one or more observed covariates. This aids, first, in understanding the magnitude of confounding required to change an answer by restating it in terms of observed covariates, for which we have stronger intuitions regarding the strength of relationship with treatment and outcome. Further, in some cases users may employ their domain knowledge and information about treatment assignment to argue that unobserved confounding is not likely to explain "k times more of the treatment assignment and outcome" than a given observable. If confounding of such strength would not change the conclusion, this can be compelling evidence

for the credibility of the research conclusion. Or, if one finds that even an optimistic assumption would still allow enough confounding to change the research conclusions, we know the original conclusion would be difficult to satisfactorily defend.

Table 1: Sensitivity summary statistics from Cinelli and Hazlett (2020)

Statistic	Interpretation
$R^2_{Y\sim D\|\mathbf{X}}$	The partial R^2 of the treatment with the outcome: In the extreme case where confounders explain 100% of the residual variance of the outcome, this tells us how much of the residual variance in the treatment such confounders would need to explain to bring the estimated effect down to zero.
RV	The robustness value: Confounding that explains at least RV% of residual variance in the treatment and in the outcome would reduce the implied estimate to zero. If both $R^2_{Y\sim Z\|\boldsymbol{X},D}$ and $R^2_{D\sim Z\|\boldsymbol{X}}$ are less than the RV, confounding is not sufficient to eliminate the effect.
$RV_{q,\alpha}$	The amount of confounding required such that an effect estimate reduced by the fraction q (e.g. 50%) would fall just within the confidence interval defined by α, i.e. the 95% confidence interval for $\alpha = 0.05$.

We employ these tools using the `sensemakr` package for `R`. Replication code for our analyses can be found at https://dataverse.harvard.edu/dataverse/jop. Additional examples, tutorials, and resources can be found in Appendix C.

3 Support for the FARC peace agreement

We now illustrate these tools by applying them to examine different explanations for public support of the 2016 peace deal with the FARC in Colombia. In October 2016, Colombians voted in a referendum on a peace agreement with the FARC, a leftist guerilla group. The peace deal was ultimately rejected, but given the immense variation in municipality-level vote share in favor of the deal (ranging from 19% to 96% in towns with at least 1,000 voters), many have sought to explain levels of support for the deal. Scholarship has coalesced around two explanations: exposure to FARC violence and political affiliation with President Santos (see Appendix A). As is often the case for substantively important questions in political science, the "treatment" of interest (exposure to violence or political affiliation) cannot be assigned randomly, nor can we argue that it is as-if random conditional on some set of observables X. Nevertheless, previous papers on this topic have relied upon various covariate adjustment approaches to address confounding (e.g. regression, matching, and weighting). Unfortunately,

none of these can hope to argue for an absence of unobserved confounders, so the results must be understood as potentially biased.[2]

3.1 Assessing evidence for the effect of exposure to violence

We consider first a naive, direct comparison by regressing $Deaths_{i,2011-2015}$, the number of deaths in municipality i committed by the FARC between 2011 and 2015, on the proportion voting "Yes" in municipality i. Table 2 presents results for such a regression (Model 1), together with the sensitivity quantities. The coefficient of 1.45 ($p < 0.001$) on violence in 2011-2015 implies that with each additional observed death, we expect to see a 1.45 percentage point higher level support for the FARC peace deal. In usual research practice, despite an awareness that confounders may have generated an unknown amount and direction of bias, such a result is typically communicated as "suggestive evidence" that exposure to violence causes higher support for peace.

Table 2: Augmented regression results for violence

Outcome: *Vote for peace deal*

Treatment:	Est.	SE	t-stat	$R^2_{Y\sim D\|\mathbf{X}}$	RV	$RV_{\alpha=0.05}$	df
1. Deaths 2011-2015	1.45	0.30	4.90	2.1%	14%	8.4%	1121
2. Deaths 2011-2015	0.61	0.29	2.11	0.4%	6.1%	0.4%	1115

The sensitivity quantities added to Table 1 quickly characterize how strong confounding would have to be to alter our conclusion. The robustness value (RV) tells us that confounding that explains at least 14% of the residual variance in both violence and support for peace would be enough to eliminate this effect entirely. Recalling that taking the square root of an R^2 allows interpreation on the usual correlation scale, this means a hypothetical confounder with a (residual) correlation of 0.34 to both the treatment (violence) and outcome (support for the peace deal) would be sufficient to explain away the entire result. Similarly, the $RV_{\alpha=0.05}$ value tells us that confounding explaining 8.4% of residual variance in violence and support would reduce the estimate to the edge of statistical insignificance. Finally, the $R^2_{Y\sim D|\mathbf{X}}$ tells us that if an unobserved confounder explains 100% of the remaining outcome variation, such a confounder

[2]Concretely, one troubling example of a potential confounder we cannot observe is "latent sympathy for the FARC". For example, suppose that those more sympathetic to the FARC also tend to be more supportive of one party or leader and more supportive of the deal, while perhaps living in areas that the FARC refrain from attacking. Such "common cause confounding" would confound both the violence and political affiliation accounts in the observed directions.

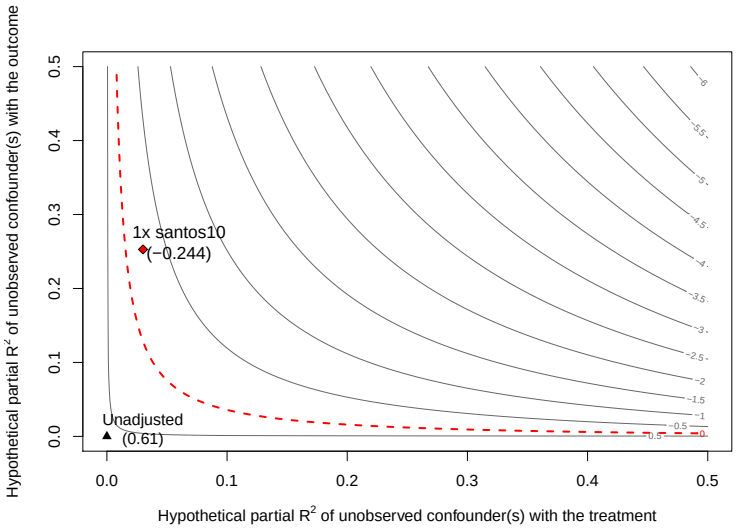
would have to explain only 2.1% of the residual variation in the violence treatment in order to reduce the estimated effect to zero.

In practice, researchers typically include control variables in hopes of mitigating uncontrolled confounding. Model 2 includes a number of potentially worrying observed confounders (see Appendix D for details). The resulting coefficient estimate is 0.61 (t=2.11). While this may pass the bar for publication as suggestive (and statistically significant) evidence in some settings, we emphasize that the statistical significance of a result alone says nothing about vulnerability in the face of unobserved confounding. A confounder explaining only 6.1% of the residual variation in both violence and support for peace would eliminate the effect; a confounder explaining only 0.4% of both would reduce the effect to the boundary of insignificance at the $\alpha = 0.05$ level. We expect readers would not have trouble coming up with confounders that arguably relate to treatment and outcome at least as strongly, such as sympathy for the FARC.

Beyond these summary statistics, a more complete picture entails visualizing the bias as we vary the strength of postulated confounding (Figure 1). These plots also visualize the results of benchmarking examples, showing how bias would be bounded subject to assumptions that relate the strength of unobserved confounding to the explanatory power of one or more observables. For example, let us assume that *political affiliation* is "stronger" than confounding, in the sense that it explains a greater share of both the outcome and treatment than can true confounding. We proxy for political affiliation using vote share for Santos in 2010, to ensure it is pre-treatment with respect to 2011–2015 violence. While assuming such a limitation on confounding may already be indefensible—it is hard to argue that *political affiliation* explains more remaining variation in exposure to violence than all other confounders could—it already admits more than enough confounding to change the sign of the estimate. Specifically, the point in Figure 1 marked "1x santos10" indicates what the adjusted estimate would be had confounding of this strength been present, bringing the original estimate (+0.61) to the opposite side of zero (-0.24).

Finally, sensitivity analyses can aid readers and reviewers even when authors did not originally provide these analyses. Using reported results in Tellez (2019) and Pechenkina and Gamboa (2019), we find an RV of 4.7% and 9.7% and an $RV_{\alpha=0.05}$ of just 1.8% and 4%, respectively, reflecting sensitivity values similar to our own (see Appendix E for details).

Figure 1: Contour plot showing sensitivity to hypothesized confounding.



3.2 Assessing evidence for the effect of political affiliation

We estimate coefficients in a model regressing the proportion voting "Yes" in municipality i on the municipal vote share for President Santos (2014), including controls for potential confounders (see Appendix D for details). Table 3 shows augmented regression results. The estimated effect of Santos vote share on support for peace (0.67) is positive and statistically significant. Vote share for Santos in 2014 explains 59% ($R^2_{Y\sim D|\mathbf{X}}$) of the residual variation in support for peace, meaning that even confounding that explains 100% of the residual variation in the outcome would need to explain 59% of the residual variation in vote share for Santos in order to eliminate the estimated effect. Confounding that explains less than 68% (RV) of both vote share for Santos and support for peace would not be sufficient to eliminate the effect.[3] For the 95% confidence interval to just include zero, confounding would have to explain 66% of residual variance in both treatment and outcome ($RV_{\alpha=0.05}$).[4] In short, very strong confounding would be required to overturn these results, as we discuss in context below.

Again, where other studies employed OLS we can determine how sensitive their results would be as well. In Krause (2017), the coefficient for Santos 2014 vote share in a similar model is 0.62, close to our estimate. The t-statistic of 45 together with 1,069 residual degrees of freedom

[3]Recall that an R^2 is just the square of r, the correlation coefficient. Thus, a partial R^2 of 0.68 corresponds to a correlation of $\sqrt{68\%} \approx 82\%$ after accounting for the other covariates—an extremely high correlation by any standard.

[4]While it appears that a similar if coarser conclusion could be drawn by simply comparing the t-statistics of the models, this is only because the sample sizes are similar across models here. More generally t-statistics and p-values cannot reflect how strong confounding must be to alter our conclusions without adjustment based on the degrees of freedom. For example, a coefficient with a t-statistic of 10 and only 200 degrees of freedom has an RV of 50%; however with one million degrees of freedom, the same t-statistics correspond to an RV below 1.0%.

Table 3: Augmented regression results for political affiliation

Outcome: *Vote for peace deal*

Treatment:	Est.	SE	t-stat	$R^2_{Y\sim D\|\mathbf{X}}$	RV	$RV_{\alpha=0.05}$	df
3. Santos 2014 vote share	0.67	0.02	37.5	59%	68%	66%	983

would produce an RV of 72%, also close to our own estimate of 68% (Appendix E). We discuss further analyses, including benchmarking, on the political affiliation hypothesis in Appendix F.

4 Discussion

Whereas asking whether a study is *entirely* free of confounding can be useful in drawing attention to the threat posed by confounding, the central concern in observational research is not whether there is *any* confounding, but whether the research conclusion might have been substantively affected by it. In any setting where arguments for exactly zero confounding may fail, we argue that more can be learned and transparently conveyed by reporting how much confounding it would take to substantively alter a research conclusion, and by using tools such as benchmarking to aid in debating whether such confounding can or cannot be readily ruled out.

The example and tools shown here illustrate how a sensitivity-based framework, if widely adopted, could improve how we conduct, communicate, and evaluate observational research. First, this approach suggests standards for empirical research seeking to make causal claims. Summary sensitivity quantities reported in augmented regression tables provide readily interpretable information about sensitivity to unobserved confounding. Here, sensitivity analyses have helped to determine that conclusions regarding the role of exposure to violence are currently too fragile to hold in high confidence. Even under the most favorable model used, confounding that explains just 6.1% of the residual variance of exposure to violence and support for peace would eliminate the result entirely, and a confounder explaining just 0.4% would reduce it below conventional levels of statistical significance. For the political affiliation hypothesis, however, a confounder explaining 100% of the residual variation in support for peace would have to explain 59% of the residual variation in political affiliation to alter our conclusions.

Second, these tools can improve how we judge the credibility and value of research projects by (i) providing readers a way of assessing how susceptible results are to confounding and (ii) encouraging critics to elaborate their concerns about confounding by arguing for specific con-

founders that they believe are strong enough to make a difference. We must remember that the large robustness value for political affiliation does not rule out the possibility that confounding has altered our conclusion. Yet, given what we know about the degree of confounding required to alter our result, a colleague or reviewer cannot suggest "just any confounder" would be sufficient to meaningfully change our conclusion. Rather, those suggesting a particular confounder are obligated to argue that it could plausibly explain the amount of variation in treatment and outcome required to alter the results.

Ultimately, we hope these tools help to bring about a change in how we value empirical projects when arguments about the absence of confounding are fallible: A paper is not to be judged by whether it convinced us that the design leaves zero confounding, but rather by how it informs our understanding of results under degrees of confounding that may plausibly exist.

References

Cinelli, Carlos and Chad Hazlett. 2020. "Making Sense of Sensitivity: Extending Omitted Variable Bias." *Journal of the Royal Statistical Society, Series B* 82(1):39–67.

Cornfield, Jerome, William Haenszel, E Cuyler Hammond, Abraham M Lilienfeld, Michael B Shimkin and Ernst L Wynder. 1959. "Smoking and lung cancer: recent evidence and a discussion of some questions." *Journal of National Cancer Institute* (23):173–203.

Krause, Dino. 2017. "Who wants peace?- the role of exposure to violence in explaining public support for negotiated agreements: A quantitative analysis of the Colombian peace agreement referendum in 2016." Unpublished Master's Thesis, Uppsala University.

Matanock, Aila and Natalia Garbiras-Díaz. 2018. "Considering Concessions: A Survey Experiment on the Colombian Peace Process." *Conflict Management & Peace Science* 35(6):637–655.

Pechenkina, Anna O. and Laura Gamboa. 2019. "Who Undermines the Peace at the Ballot Box? The Case of Colombia." *Terrorism and Political Violence* pp. 1–21.

Tellez, Juan Fernando. 2019. "Worlds Apart: Conflict Exposure and Preferences for Peace." *Journal of Conflict Resolution* 63(4):1053–1076.