Qualitative Imputation of Missing Potential Outcomes

Alexander Coppock and Dipin Kaur*

July 23, 2020

Abstract

We propose a framework for meta-analysis of qualitative causal inferences. We integrate qualitative counterfactual inquiry with an approach from the quantitative causal inference literature for reasoning about missing information called extreme value bounds. Under the Neyman-Rubin model, units are endowed with potential outcomes, or responses that units would express depending on the level of some treatment. The goal of qualitative counterfactual analysis is to use the observed outcome and auxiliary information to infer what would have happened had the treatment been set to a different level. In other words, the analyst attempts to *impute* missing potential outcomes. Imputation is hard and when it fails, we can fill in the missing potential outcomes with best- and worst-case scenarios. We show how the resulting extreme value bounds represent fundamental uncertainty and how the imputation of missing potential outcomes can shrink that uncertainty in a structured manner. We provide an application of our approach to 121 cases that could have experienced transitional truth commissions, 16 of which did. Prior to any analysis, the extreme value bounds around the average treatment effect are 100 percentage points wide; incorporating qualitative beliefs about counterfactuals shrinks the width of these bounds to approximately 40 points.

[*] Alexander Coppock is Assistant Professor of Political Science, Yale University. Dipin Kaur is a Ph.D. Candidate in the Department of Political Science, Yale University. The authors thank Peter Aronow, Angele Delavoye, Morgan Galloway, Michael Goldfein, Matthew Graham, Cleo O'Brien-Udry, Jason Seawright, Drew Stommes, Libby Wood, and Lauren Young for helpful conversations and discussions.

A special promise of qualitative counterfactual inquiry is that deep case knowledge can generate informed guesses of *what would have happened if things had been different.* We observe realized outcomes but by definition we can never observe counterfactual outcomes – they are counter-to-fact. Therefore, a causal inference in a single case amounts to a claim about the value of a missing potential outome. For the purposes of this paper, we call such claims "imputations."

We acknowledge from the outset that counterfactual analysis is not usually the only purpose of qualitative research, but it is often at least one of the purposes, and this article is about those purposes only. Our goal is to show how a systematic aggregation of counterfactual guesses can shrink fundamental uncertainty about average causal effects in a principled manner. Extreme value bounds (Manski, 1999) represent the logical range of average causal effects that are consistent with the world as we observe it. The bounds start out quite wide, but we show in this article how the qualitative imputation of missing potential outcomes can tighten the bounds substantially.

Our approach shares much in common with those of Seawright (2016), Glynn and Ichino (2015) and Humphreys and Jacobs (2015), each of whom incorporate qualitative and quantitative information to produce better *estimates* of average causal effects. By contrast, we leave entirely to the side the question of when and how qualitative researchers should draw causal inferences in a single case. Doing so is generally quite difficult, but obviously not impossible. Qualitative methodologists have developed a battery of approaches to single-case causal inference, and different cases may call out for different methods (see, e.g., Goertz and Mahoney, 2012; George and Bennett, 2005; Bennett and Checkel, 2014; Ragin, 2014). We seek to incorporate single-case causal inferences (no matter their specific methodological provenance) within a synthesized framework, something akin to a meta-analysis for qualitative inquiry. Our approach leaves open the possibility that imputation may fail in a particular case – the researcher may conclude that the existing information is too thin to merit a confident counterfactual guess and may decide to leave a counterfactual outcome altogether unimputed. Alternatively, the analyst may express an *uncertain* guess (in the form of a probability statement) about counterfactual outcomes; the resulting meta-analysis incorporates that uncertainty.

Much recent work has sought to combine quantitative and qualitative methodologies for causal inference (Gary King, 1994; Brady and Collier, 2010; Mahoney, 2010; Humphreys and Jacobs, 2015; Glynn and Ichino, 2015), often through a potential outcomes framework (Seawright, 2016; Morgan and Winship, 2014; Acharya et al., 2016; Abadie et al., 2015; Imai et al., 2011; Pearl, 2009). However, counterfactual analysis has a long tradition among qualitative methodologists as well (Hume, 1748; Lewis, 1979, 1973; Woodward, 2005). Three examples from recent empirical scholarship illustrate the diverse range of approaches to counterfactual analysis used by qualitative researchers.

Harvey (2012) evaluates the conventional wisdom that the US decision to invade Iraq in 2003 was a product of neoconservative ideology, internal delusions and grand strategies unique to President George W. Bush and his national security team. Here, the Bush presidency is the treatment condition, a counterfactual Al Gore presidency is the control condition, and the outcome of interest is the Iraq War. The world revealed the treatment outcome: $Y(Bush) = War$. Harvey leverages a comparative counterfactual framework to impute the unobserved control outcome, $Y(Gore)$. To this end, he analyzes "facts and evidence derived from a careful (and complete) review of the relevant historical record" (Harvey, 2012) including analyses of interviews, speeches and public statements. The author finds little support in favor of the generally-accepted counterfactual that a Gore administration would have remained at peace. Instead, he concludes that the onset of the Iraq war was a product of many key decisions and entrenched misconceptions that constituted the path dependent sequence of moves pushing the US-UK coalition closer to war, thereby implying that $Y(\text{Gore}) = \text{War}$ as well.

Haber and Menaldo (2011) operationalize explicitly specified counterfactuals to evaluate the resource curse hypothesis that countries with natural resource dependency are less likely to be democratic. The treatment here is the country's reliance on natural resources and the outcome of interest is the regime type. To evaluate this relationship, Haber and Menaldo specify the counterfactual path that a resource-reliant country's regime type would have followed in the absence of those resources on the basis of the path followed by non-resource reliant countries in its geographic region. They then compare the counterfactual path to the actual one, to see whether any divergence between the two paths of political change is correlated with increases in resource reliance. They conclude, contrary to conventional wisdom, that an increase in natural resource reliance does not promote authoritarianism. In other words, they claim that $Y(\text{Resources}) = Y(\text{No Resources})$.

Lastly, Lebow and Stein (1996) assess the counterfactual claim made by Khrushchev that had the Soviet Union not deployed missiles in Cuba triggering the Cuban Missile Crisis, the United States would have invaded the island. In this case, the treatment is the Soviet deployment of missiles and the outcome is American invasion into Cuba. While history has revealed the treated outcome, Lebow and Stein use recently uncovered evidence to impute the unobserved untreated outcome. They show that even before the missile deployment, no influential members of the Kennedy administration wanted to attack Cuba. The option had been considered but decisively rejected. Kennedy and Secretary of Defense McNamara were impressed by the level of Cuban popular support for Fidel Castro and the ability of the Cuban militia to overwhelm the invasion by force. Hence, they were deterred by the revised intelligence estimates which indicated that a successful invasion would have required massive US forces to remain in an occupational role for an indefinite period. Based on these costs, they resolved to not attack unless there were dramatic political changes inside Cuba. They conclude that Khrushchev and his advisors were unambiguously wrong: Soviet missiles were not necessary to prevent an American attack. If anything, the missile deployment only increased foreign policy and domestic political pressure for President Kennedy to invade. In this case, $Y(\text{Missles}) = \text{No invasion}$; Lebow and Stein (1996) claim $Y(\text{No missles}) = \text{No invasion}$ as well.

The foregoing examples demonstrate that our approach is emphatically mixed-methods. Each of the three causal inferences relies on qualitative, case-specific data to generate counterfactual claims. Examples 1 and 3 used process tracing while example 2 relied on case comparison. The beauty of describing causal effects in terms of counterfactuals is that we can express the research findings in a common language, regardless of the precise qualtitative inference procedure used to generate the findings. Once the causal effects in a literature have been expressed in that common language, they can be aggregated and synthesized.

Our approach combines counterfactual imputation with extreme value bounds (explained in detail in the next section). The main purpose of the method is to structure and characterize the uncertainty surrounding average causal effects. The bounds reflect fundamental uncertainty because they are a function of what we know for sure (the data revealed by the world), what we think we know (the inferences we draw in some or all cases), and what *we know* we do not know.

Describing counterfactuals using potential outcomes

Under the Neyman-Rubin causal model (Neyman, 1923; Rubin, 1974), units are endowed with a set of potential outcomes, only one of which they reveal depending on the realization of exposure to causal agents. In the most basic case, a unit i has only two potential outcomes, $Y_i(1)$ and $Y_i(0)$, which correspond to the treated and untreated potential outcomes. This setup embeds an assumption of noninterference, or the assumption that unit i does not have potential outcomes beyond $Y_i(1)$ and $Y_i(0)$ that might depend on the treatment assignments of other units.[1] The realized treatment d_i then "reveals" the observed outcome Y_i via the "switching" equation: $Y_i = Y_i(1)d_i + Y_i(0)(1 - d_i)$. If treated, unit i reveals $Y_i(1)$ and if untreated, it reveals $Y_i(0)$.

The fact that we can never observe a unit in both its treated and untreated state has been famously dubbed the Fundamental Problem of Causal Inference (Holland, 1986). In Table 1, we see a treated unit ($d_i = 1$) and its observed outcome ($Y_i = 1$). We know $Y_i(1)$ because it is equal to the revealed outcome Y_i. The untreated potential outcome $Y_i(0)$ is "missing" in sense that we do not know its value. The goal of counterfactual analysis is to fill in the missing value with a (hopefully very well-educated) guess. Since the outcome in this example is binary, this amounts to filling in the question mark with either a zero or a one.

Table 1: Causal inference for a single unit

d_i	Y_i	$Y_i(0)$	$Y_i(1)$
1	1	?	1

One of the analytic tasks of qualitative research is to understand the separate impacts of the many causal factors that explain outcomes in a single case. Moreover, qualitative researchers often

[1] Sometimes noninterefence is referred to as SUTVA, or the "stable unit treatment value assumption, " but SUTVA encompasses an additional assumption beyond noninterference, rendering its invocation slightly ambiguous.

consider the mechanisms by which treatments affect outcomes. Both questions – *which* factors matter and *why* – are important, complicated, and difficult to answer. We will focus on a tiny slice of that analytic task: understanding the impact of a single causal factor on a single outcome, inclusive of any and all mechanisms that may be at play in a particular case. Firstly, that means we will restrict our attention to "effects-of-causes" questions as opposed to "causes-of-effects" questions (Goertz and Mahoney, 2012).[2] Understanding the impact of just one treatment on just one outcome is hard enough; an exhaustive accounting of all of the causes of an outcome may be more or less out of reach in most cases. Secondly, we are after the total effect of the treatment on the outcome rather than the portions of the total effect that can be attributed to various intermediate processes. For a discussion of the extreme difficulty inherent in studying mechanisms (mediators), even when treatments are randomly allocated to subjects, see Bullock et al. (2010) or Gerring (2010).

Causal inference for a single unit requires the researcher to *impute* the missing potential outcome. Using case knowledge, information about individual actors' incentives, institutional arrangements, temporal variation, and logic, qualitative researchers can make a guess about what would have happened the treatment had been set to a different level. The uncertainty attending to that guess can also be qualitatively expressed. For some units, this task is easy. For others, it is much harder. Qualitative methodologists have developed a suite of approaches for determining whether the available qualitative data are sufficient to license a causal inference (George and Bennett, 2005; Gerring, 2006; Mahoney, 2010; Collier, 2011; Goertz and Mahoney, 2012; Mahoney, 2012; Rohlfing, 2012; Fairfield, 2013; Bennett and Checkel, 2014; Ragin, 2014; Beach and Pedersen, 2016). In our empirical application, we rely in large part on (our reading of) the qualitative inferences drawn by other researchers, each of whom has made choices among the variety of methods available to them. As our application below will show, sometimes no approach (qualitative, quantitative, or otherwise) is sufficient for causal inference and we are forced to admit ignorance of causal effects. We view the ability to incorporate the *lack* of knowledge about counterfactuals as a major strength of our procedure.

[2] The "effects-of-causes" and "causes-of-effects" terminology is somewhat awkward. An "effect-of-a-cause" question is about the difference in outcome Y_i depending on the level of the cause d_i. The "causes-of-effects" questions should really be described as "causes-of-outcomes" outcomes, since the goal of such inquiries is to enumerate the causes that eventuate in the observed level of outcome Y_i.

Our goal will be to summarize qualitative inferences about individual-level causal effects for a set of N units. In particular, we aim to place bounds around the average treatment effect (ATE) for these N units: $ATE \equiv \frac{\sum_i^N Y_i(1)-Y_i(0)}{N}$. The ATE is a common target of inference in quantitative research but less so in qualitative work (Goertz and Mahoney, 2012). One might reasonably argue that a chief advantage of qualitative methods is that they are addressed to targets that are far more subtle than a simple average over possibly very heterogeneous cases. While we will mainly focus on the ATE in this paper, our procedure could easily be extended to many other estimands, including conditional average treatment effects (the ATE conditional on membership in a particular group of units), or the average treatment effect on the treated (ATT, or the ATE among those units that come to be treated). In fact, nothing about our technique limits researchers to studying average effects because the bounding approach can be extended to any summary of the full joint distribution of potential outcomes in a population. If a question can be stated in terms of potential outcomes, we can use qualitative imputation of counterfactuals in combination with bounds to characterize both an answer to the question and our uncertainty.

The procedure

In this section, we describe the procedure in the case binary outcomes and only one kind of uncertainty: either the analyst is certain of the counterfactual outcome or they are not. In the extensions to follow, we consider non-binary outcomes and a second kind of uncertainty, namely, the expression of counterfactual beliefs as probabilities.

Extreme value bounds (Manski, 1999) are the logical bounds around the ATE.[3] Consider a setting with a binary outcome and a binary treatment. In the "best" case, the outcome for everyone in the treatment group is "1" and the outcome for everyone in the control group is "0." In this scenario, the ATE is +100 percentage points. By the same logic, in the worst case scenario, the ATE is -100 percentage points. Before any data are collected, the extreme value bounds are 200

[3] Bounds are often used when outcome data are missing. See Gerber and Green (2012, chp. 7) for an accessible introduction to bounding approaches for attrition. Aronow et al. (2017) applies bounds to the case of researcher-induced attrition (dropping subjects who fail a manipulation check) and Coppock (2018) applies bounds to the analysis of experiments in which researchers condition on one post-treatment variable to study the effects of treatment on a second post-treatment variable.

percentage points wide, which correctly characterizes our utter ignorance of the ATE.

Once the data are collected, we observe each unit in either its treated or untreated state and we observe the associated potential outcome. In the control group we observe $Y_i(0)$ but not $Y_i(1)$, and in the treatment group we observe $Y_i(1)$ but not $Y_i(0)$. If we now impute the best case and worst case scenarios, we only have to impute *half* of the potential outcomes because the world has revealed the other half. Once the data are collected, the extreme value bounds have shrunk from 200 points wide to 100 percentage points wide. These bounds represent – before the inclusion of any priors, qualitative information, or other expertise – the uncertainty attending to the ATE. This uncertainty is not due to the sampling or assignment procedures, but instead to the fact that we only observe half the data and we are uncertain about the other half.

In order to shrink the width of the bounds around our uncertainity, we repeat the following two steps, adding possibly stronger assumptions each time, until the reservoir of qualitative case knowledge on the topic has been drained.[4]

1. Impute missing potential outcomes using qualitative case and process knowledge

2. Recompute extreme value bounds

A toy example

Consider a population of $N = 10$ units, seven of whom have been treated and three of whom have not. We observe the revealed (binary) outcome for all ten units. The outcome for three of the treated units and one of the untreated units is 1; the outcome is equal to zero for the remaining units. Because the treated and untreated units may differ in both observed and unobserved ways, we cannot simply compare treated to untreated. Instead, we will impute best and worst case scenarios for the unknown potential outcomes. Before adding any qualitative information, the bounds on the ATE extend from -40 percentage points to 60 percentage points. Table 2 presents the table of potential outcomes as it proceeds through three rounds of imputation. Unknown and unimputed

[4] In the toy example and empirical exercise below, we consider whole "batches" of imputations at a time as a way to structure the successive accumulation of evidence. The order in which particular cases are imputed does not matter for the final width of the bounds, unless later imputations depend on the choices made in earlier cases.

potential outcomes are represented with question marks and imputed outcomes are shown in bold red.

The 4th and 5th columns describe the imputation of the five "easy" cases. These are scenarios in which the untreated outcomes for units 1, 2, and 5 are obviously (to the researcher) 1, 1, and 0, respectively. Similarly, the treated outcomes of units 8 and 9 are 1 and 0. These cases are "easy" in the sense that it is clear to the researcher that the treatment could not have had an effect on the outcome, perhaps because the outcome was clearly the consequence of a complex set of factors that exclude the treatment under consideration. These five imputations have shrunk the extreme value bounds considerably, and they now reach from -20 points to 30 points.

Suppose the researcher reasons next that because the treatment should have, if anything, had a positive effect for units 6 and 7. Since the revealed treated outcomes were 0 for both units, the research concludes that the treatment must have have no effect on these units and imputes 0's for the untreated outcomes as well. This imputation shrinks the bounds to $[0, 30]$.

As a final step, the researcher invests heavily in the remaining "hard cases," units 3, 4, and 10. Suppose that the efforts are only partially rewarded. In cases 3 and 4, the researcher concludes that the treatment had a positive effect, and so imputes a 0 for the untreated outcomes in those cases. But in case 10, the empirical record is too thin and the causal story too murky to make a confident call. The resulting bounds $[20, 30]$ correctly incorporate the remaining uncertainty in the researchers' summary of beliefs about causal effects.

Figure 1 shows the width of the extreme value bounds at each step in the process. The incorporation of both quantitative and qualitative information reduces the width of the bounds from 200 points to 10 points, corresponding to a dramatic reduction in fundamental uncertainty.

Extensions

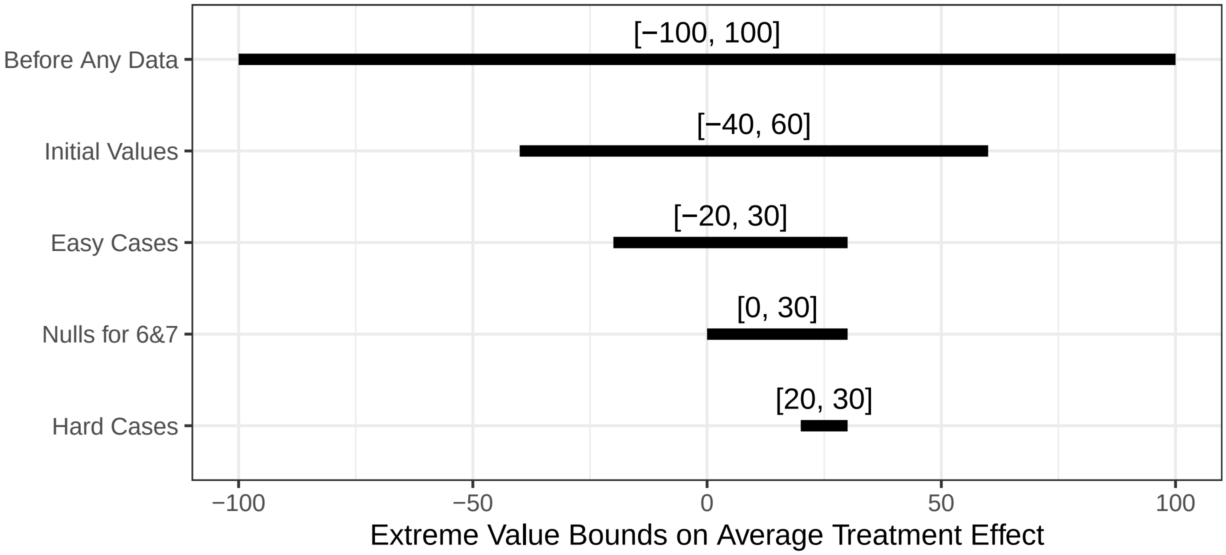
Here we consider two extensions of the basic procedure.

First, some analysts may balk at the idea of giving a definitive counterfactual imputation, equivalent to claiming with certainty that the counterfactual outcome would have taken a particular value. High certainty claims about causal effects are of course the goal of qualitative counterfactual

Table 2: Toy application: Potential outcomes table

	Observed		Initial values		Easy cases		Nulls for 6&7		Hard cases	
	d_i	Y_i	$Y_i(0)$	$Y_i(1)$	$Y_i(0)$	$Y_i(1)$	$Y_i(0)$	$Y_i(1)$	$Y_i(0)$	$Y_i(1)$
1	1	1	?	1	**1**	1	1	1	1	1
2	1	1	?	1	**1**	1	1	1	1	1
3	1	1	?	1	?	1	?	1	**0**	1
4	1	1	?	1	?	1	?	1	**0**	1
5	1	0	?	0	**0**	0	0	0	0	0
6	1	0	?	0	?	0	**0**	0	0	0
7	1	0	?	0	?	0	**0**	0	0	0
8	0	1	1	?	1	**1**	1	1	1	1
9	0	0	0	?	0	**0**	0	0	0	0
10	0	0	0	?	0	?	0	?	0	?
	EV Bounds:		[-40, 60]		[-20, 30]		[0, 30]		[20, 30]	

Figure 1: Toy application: Extreme value bounds



inference, but the available evidentiary basis (qualitative or quantitative) might be too weak. In such cases, we may be able to express a guess about counterfactuals in terms of a probability rather than a full imputation.

We can incorporate probablistic beliefs with a slight elaboration of the bounding procedure. We divide unknown potential outcomes into three classes: those we can impute with certainty, those we cannot impute at all, and those for whom we can state the probability of the binary outcome taking on the value 1. If there are k potential outcomes in this third class, then we have to consider 2^k possible sets of potential outcomes, each of which would occur according to the joint probability specified by the analyst.[5] In each of the 2^k possibilities, we can compute extreme value bounds. The "point estimate" for the bounds is equal to its expectation, or the probability-weighted sum over all 2^k possibilities. We can then express our uncertainty about the location of the bounds as the standard deviation of the set of bounds that would occur according to analysts beliefs about the probabilities of the counterfactual outcomes. In practice, this set may be too numerous to enumerate fully, in which case we can draw an arbitrarily large random sample from the full set of 2^k possibilities.

This extension highlights the difference between two sources of uncertainty: ignorance of the missing potential outcome and beliefs about the probability that the missing potential outcome equals 1. The first is about fundamental uncertainty. Because of the fundamental problem of causal inference, we do not know what would have happened had the treatment been set to a different level. The second is about a specific belief on the basis of case knowledge about the probabilities of counterfactual outcomes. To see this, suppose that an analyst claims that the probability of the missing potential outcome is 0.5. This "coin flip" imputation is importantly different from making no imputation at all because the resulting extreme value bounds will be *tighter* than if the outcome were left unimputed altogether. Stated differently, claiming that the missing potential outcome is equally likely to be a 1 as it is to be a 0 requires more qualitative information than leaving the

[5] If the probabilistic beliefs about counterfactuals are independent across units, then the joint probability of each set of potential outcomes is equal to the product of the probability that each of the k counterfactual outcomes would take on the value expressed in that particular set. If these guesses are dependent (an analyst might think, either both of these units would have expressed a 1 or neither would have), then the analyst can articulate specific beliefs about joint probabilities.

outcome unimputed.

We can also extend our framework beyond binary outcomes. While many of the most important outcomes are binary, others many be continuous or quasi-continuous. Suppose the logical extrema of the outcome variable are Y^{MAX} and Y^{MIN}. The width of the bounds can be computed as $\frac{[(Y^{MAX}-(Y^{MIN})*N_{unimputed}}{N}$, where $N_{unimputed}$ is the number of units for which we cannot make a confident counterfactual prediction.

We can also substitute the values of Y^{MAX} and Y^{MIN} for the unknown potential outcomes in order to calculate the maximum and minimum possible values for the average treated ($\overline{Y_1}$) and untreated ($\overline{Y_0}$) potential outcomes. These are

$$\begin{aligned}
\overline{Y_1^{MAX}} &= \frac{\sum_1^m Y_i + (N-m) * Y^{MAX}}{N} \\
\overline{Y_1^{MIN}} &= \frac{\sum_1^m Y_i + (N-m) * Y^{MIN}}{N} \\
\overline{Y_0^{MAX}} &= \frac{\sum_{m+1}^N Y_i + (m) * Y^{MAX}}{N} \\
\overline{Y_0^{MIN}} &= \frac{\sum_{m+1}^N Y_i + (m) * Y^{MIN}}{N}
\end{aligned}$$

, where the first m of N units are treated and the remainder are untreated. The ATE cannot be larger than $\overline{Y_1^{MAX}} - \overline{Y_0^{MIN}}$ nor can it be smaller than $\overline{Y_1^{MIN}} - \overline{Y_0^{MAX}}$.

Imputation with continuous outcomes works in the same way as with binary outcomes. Using the observed data and qualitative information, the researcher makes an informed guess as to what the counterfactual outcome would have been had the treatment been set to the other level. Doing so may be more challenging in the continuous outcome case, since the analyst is required to pick one value out of a range of values, rather than make a (comparatively) easier call about a binary state. A middle ground between affirmitively claiming a particular value for the unobserved potential outcome and leaving the outcome altogether unimputed is to redefine – in particular cases – what the best and worst cases are for the counterfactual outcome. For example, imagine units are countries and the outcome is a 1 - 7 Freedom House score (the higher the score, the less free the country). Suppose the observed outcome for a treated country is 2, and the analyst's goal is to impute its untreated outcome. The analyst believes for good reasons that the treatment improved outcomes for the country (so the score would have been higher under control) but is unsure how much higher. For that unit only, we could define Y^{MAX} as 3 and Y^{MIN} as 2. The resulting extreme bounds around the ATE will be tighter than if we brought no information to bear at all.

Application to the average effect of truth commissions

In this section, we apply our procedure to the study of transitional truth commissions (TTCs), an area where a considerable amount of case-specific qualitative information has been generated by researchers within and without the academy. Studying the average effect of TTCs using standard quantative methods like matching or regression with controls is probably too difficult because those places that come to be treated differ in so many ways (both observed and unobserved) from those places that are untreated. In this setting, a claim that one has both measured and correctly adjusted for all possible confounders amounts to a leap of faith that is in our view not worth taking. We understand that opinions will differ on this point, but we want to note that it was our skepticism of the applicability of regression-like approaches that motivated the development of our procedure in the first place.

Our first task was to define the universe of cases that were *eligible* for a TTC. Clearly, all cases that received a TTC were eligible; the difficulty was finding those cases that could have but did not experience a TTC. We obtained information on two types of political transitions: transition from civil war and transitions to democratic rule. We limited our search to the period from 1980 to 2008 because the first completed truth commission was established in 1983 and we need to allow some time to elapse after a truth commission is possible in order for the world to reveal outcomes. We identify civil conflict transition cases eligible for truth commissions from the Kreutz (2010) Uppsala Conflict Data Program's (UCDP) Conflict Termination Dataset and democratic transition cases from the Geddes et al. (2014) dataset.

Definition of treatment

Our treatment is the establishment of a TTC. We use the broad definition of truth commissions given in Hayner (2000): "officially created investigative bodies that document patterns of past human rights abuse over a specified period of time." Of course, precisely which bodies meet this (or similar) definition has been the subject of much debate. Some lists of TTCs are relatively expansive (USIP, 2011*b*; Olsen, Payne and Reiter, 2010; Hayner, 2006) and others are relatively conservative (Dancy et al., 2010; Bakiner, 2015; Kim and Sikkink, 2010). In keeping with the more restrictive accounts, we consider a body a TTC if it (i) investigates for a limited amount of time; (ii) publishes a final report; (iii) examines a limited number of past events along with their patterns, causes and consequences; (iv) enjoys autonomy from direct intervention by political actors; and (v) remains official in character (Bakiner, 2015; Dancy et al., 2010).

Truth commissions can be further separated into two types. Bakiner (2013) differentiates "transitional" truth commissions (those that come up within the first three years of transition to peace or democracy) from "non-transitional" truth commissions (those that do not arise in the context of transition), arguing that the two types display specific dynamics and require different analytic tools. In addition, Wiebelhaus-Brahm (2009*b*) describes non-transitional commissions as "historic" truth commissions and argues that they are qualitatively different from commissions that take place during a transition. We therefore we only consider as "treated" the fifteen cases that Bakiner considers to be transitional.

The "untreated" category then includes not only cases that experience no commission whatsoever but also those that fall into a variety of other scenarios. For instance, countries that establish truth commissions many years after transition (Uruguay, South Korea, Panama, Brazil), those that announce truth commissions but do not implement them (Burundi), those created by authoritarian governments (Morocco), and those initiated in the context of limited demobilization during ongoing conflict (Colombia) are considered untreated. Also in this category are unofficial commissions created by community processes (such as Brazil), commissions of inquiry set up to investigate a singular event (such as a riot, pogrom or massacre leading to disappearances) and those set up to investigate embezzlement, fraud and similar crimes (as in Olsen, Payne, Reiter and Wiebelhaus-

Brahm (2010)). Finally, we consider 'sham' commissions (installed for window-dressing such as Sri Lanka 2010) and disbanded commissions (that are unable to complete their work such as Bolivia, Ecuador) as untreated.

In randomized experiments, subjects are sometimes assigned to be treated but they fail to take treatment. Under an exclusion restriction that the assignment itself does not affect outcome, these "noncompliers" reveal their *untreated* potential outcome. In cases that experienced commission-like bodies that do not meet our criteria for a TTC, we also assert that they, like noncompliers, reveal their untreated outcomes. Our task then is to impute their *treated* potential outcome, i.e., what would have happened had the unit experienced a full TTC according to our definition.[6]

When constructing counterfactuals, a major difficulty is specifying a "conceivable" counterfactual world that has enough detail to be both meaningful and plausible (Fearon, 1991; Tetlock and Belkin, 1996). That is, for every case that experienced a TTC, there are an infinity of ways to not experience a TTC; a similar logic holds for the untreated group. We followed protocols developed by scholars for conducting robust counterfactual thought experiments (Lebow, 2010; Tetlock and Belkin, 1996), and consequently endeavored to imagine a world that is as similar as possible to the one that occurred, but for the difference of having a TTC.

Definition of outcomes

The main reason to separate the end-of-conflict cases from the democratization cases is the definition of outcomes. One major advantage of treating this problem qualitatively is that we have no need to analyze the same outcome variable for all cases simply to maximize N. Instead, we consider the outcome that seems most appropriate to each set of cases. For the end-of-conflict cases, our outcome is the recurrence of violence whereas in the democratization cases, our outcome is the resumption of authoritarianism.

[6] For readers familiar with the argot of randomized experiments, notice that this choice deviates from the "intention-to-treat" principle, according to which units are analyzed according to their treatment *as assigned* rather than according to the treatment status *as revealed*. In an experiment, the intention-to-treat principle should be followed because it preserves symmetry across the assigned treatment and control groups. However, since our goal is the aggregation single-case causal inferences, we can be more flexible. Again under an exclusion restriction, units that were assigned a TTC but for some reason did not experience one express their untreated outcome, so our task is to impute the treated outcome.

Since we are interested in the medium-term impact of truth commissions, we record outcomes as 1 for each case where conflict or authoritarianism resumes within ten years from the termination of the conflict, and 0 otherwise. In order to ascertain the universe of cases, we obtained information on two types of political transitions that precede the establishment of truth commissions: transition from civil war and to democratic rule starting 1980 (the first completed truth commission was established in 1983) and until 2008 (allowing us ten years to observe conflict recurrence). We identify civil war transition cases and their recurrence from the Kreutz (2010) Uppsula Conflict Data Program's (UCDP) Conflict Termination Dataset. Based on this dataset, each entry is a separate war that reaches an intensity of at least 1000 cumulative battle deaths (Blattman and Miguel, 2010). Conflicts below this casualty threshold are considered minor conflicts by UCDP and are hence excluded. Further, we arrive at instances of democratic transition from various types of authoritarian regimes and their resumption from the Geddes, Wright and Frantz (2014, GWF) Autocratic Regimes Dataset. Autocracy, according to the dataset, is defined as a set of formal or informal rules for choosing leaders and policies and each entry refers to consecutive years in which the same autocratic regime has been in power in a particular country. We combine these datasets with qualitative accounts of timelines and activity within individual conflicts. Based on these selection criteria, we arrive at two datasets that include 54 end-of-conflict observations and 67 democratization observations. These datasets represent a universe of possible truth commission cases at transition, and we analyze the two entirely separately.

Imputing missing potential outcomes

We attempted to impute the missing potential outcome in 121 cases and we succeeded in making 65 imputations. We break up the imputations into four large categories. These categories represent the ease of imputation based on the depth of scholarship available on each case. However, any change in the order in which we consider individual case would not impact our final result in any way under the assumption that would make the same imputation regardless of the order in which we consider our cases. We give the main reasons for our choices here and provide short descriptions of all cases in the appendix. At the same time, our framework remains flexible to changes in particular imputations if critics disagree with our arrived-upon guesses.

Step 1: Disbanded and discredited cases

As the first step, we identify truth commissions that were either established but disbanded before completion (Bolivia in 1982, the Philippines in 1986, the Federal Republic of Yugoslavia and Nepal in 2006), or were established solely for the purposes of window dressing and political maneuvering (Sri Lanka in 2010, Croatia in 2004). These units reveal their *untreated* outcome.[7] The observed outcome ($Y_i(0)$) in each of these four cases has been 0, as conflict did not resume within ten years of transition in any of these cases. For our purposes, these cases are therefore untreated units with an observed outcome of 0.

Upon researching each individual case, we found that they were disbanded (or reduced to window-dressing) due to underfunding or the lack of political will to investigate violations. We determined that they were not left incomplete or rendered ineffective because of fears of renewed violence. For instance, in the case of Bolivia, Hayner (2000); Skaar (1999); USIP (2011*a*), find that the 1982 truth commission was unable to complete its work because of financial constraints. Before it was disbanded, however, the commission managed to document 155 cases of disappearance. Even though none of the cases were conclusively investigated, the attempt itself sparked numerous civil society debates. The ensuing public pressure eventually led political figures set aside the initial amnesty law (from the time of transition) that protected the outgoing military regime from prosecution and institute the trials against more than 50 former officials of the military government. While these trials were not based on evidence gained by the Truth Commission, but "the combination of a truth commission, trials, and private efforts at truth-finding resulted in what Human Rights Watch and others characterized as an overall positive process" (Hayner, 2000, p. 54). In light of different, though initially incomplete transitional justice attempts reinforcing each other without authoritarian backlash, it is unlikely that the truth commission, had it been carried to completion, would have catalyzed a return to military dictatorship.

[7] As an additional robustness check, we re-compute extreme value bounds under the assumption that all cases where a truth commission was set up, whether or not it was considered independent carried to completion, reveal their treated outcome (see Appendix D).

In contrast, a truth commission was also established in Sri Lanka, but only as a means to deflect international pressure to investigate human rights violations (International Crisis Group, 22 December 2011) after its clear victory against the LTTE. Even though the commission issued various recommendations and submitted a report after 18 months of investigation, it is widely denounced given its limited mandate and its failure to meet international standards of independence (Höglund and Orjuela, 2011; Yasmin Sooka, 2019). We impute that the even if independently conducted, the truth commission would not have caused a return to conflict. At the domestic level, multiple official attempts to investigate events of the civil war in the form of commissions of inquiry have had no adverse impact on rebel grievances (Lonergan, 2017). Although these efforts have been criticized for their limited scope, they serve to demonstrate that investigation into violence is not antithetical to continued stability. Instead, Tamils and Muslims displayed a strong desire to participate in renewed truth commission processes, showing up to the capital to give testimony despite turbulent conditions (Thiranagama, 2013).

These case-specific analyses lead us to believe that had a "real" truth commission been set up and completed in these cases, it would not have contributed to a resumption in conflict. We found compelling, non-security related reasons that explained why the commissions were not fully executed (a lack of political will). On this basis, we impute the treated potential outcome in each of these cases to be $Y_i(0) = Y_i(1) = 0$.

Step 2: Treated cases

Next, we turn to the treated cases, which are transitional cases that received bona fide transitional truth commissions. We identified 15 such cases (in line with Bakiner (2015). Most of these cases have been the subject of extensive scholarship, which allows us to draw on existing analyses from multiple sources. For instance, experts have catalogued data on the effects of the South African Truth and Reconciliation Commission on a variety of outcomes. Landmark studies by Gibson (2006, 2004, 2002); Wiebelhaus-Brahm (2010) deem the truth commission as key to South Africa's transition. They find that claims of South Africans being dissatisfied with their commission were largely limited to White South Africans, and more than 85 percent of Black South Africans interviewed believed that ”the commission did a reasonable job of letting families know what happened to their loved ones, of providing a true and unbiased account of the country’s history, and of ensuring that human rights abuses would not happen again” (Gibson, 2002). More importantly for our purposes, Wiebelhaus-Brahm (2010) conducted an explicit counterfactual analysis around the South African commission’s contribution to its democracy to argue the following: “a brief counterfactual suggests that the TRC did play a significant role in this regard [contribution to democratic institutions]. Imagine a South Africa in which the TRC did not exist. Perhaps the NP was able to extract a blanket amnesty as a concession for giving up power. Vigilantism would likely have exploded and whites would have fled South Africa in even larger numbers. Conversely, a South Africa in which many apartheid government officials were put on trial would seem a likely recipe for civil war. Many observers believed whites would prefer civil war to being ruled by the ANC. As it turned out, the TRC did just enough to satisfy all sides” (48). In this context, even though the government’s failure to institute timely and adequate reparations to victims immediately following the Truth Commission created renewed political tensions (Laplante and Theidon, 2007), the investigation into and reparations for violations likely prevented their repetition. We therefore interpret the South African case as one where in the absence of the TTC, the transition would have been incomplete and repression would have been likely to resume. In other words, $Y_i(0) = 1$.

On the other end of the spectrum, we consider a case like Nigeria, which faced both moral and practical dilemmas in the setup of its Human Rights Violation Investigation Commission (HRVIC). First, the President who took over after the transition to democracy was himself once the leader of the military regime that was under investigation under the truth commission’s mandate. In such a situation, “for Nigerians, the very source of the HRVIC seemed antithetical to its goals of reconciling the past, because it was set up by a former military leader who was still, either negatively or positively, engaged with different cliques of Nigeria’s corrupt and ethnicized military elite” (Nwogu, 2007). Additionally, the government faced practical resource constraints despite subsidies from major international players. Nwogu argues that “the government, while seeking the dividends of a truth commission, did not have the patience to allow for the time needed to create the appropriate and context-specific procedures, nor did they budget the necessary funds needed to accomplish a thorough job... while Osbanjo wanted to legitimate the status of his government as a transitional one, he was not particularly in favor of undertaking the challenges of a transitional government. Therefore, he used the HRVIC as a means to distance himself from his predecessors but was not vested in acting on the moral claim that a transitional government pronounces against the past" (39). In light of these issues, experts have concluded that "[t]he failure of the process derived not from any major deficiency in the work of the Commission but largely from external factors, chief amongst which was a lack of sincerity on the part of the initiating regime... there was deliberate financial strangulation in order to ensure that the Panel became a political weapon in the hands of the President against the potential contenders for the presidency in the 2003 elections" (Yusuf, 2007). In this and other similar cases of TTCs that completed their work but prioritized perverse political ends over the updating of historical record and the creation of follow-up institutions, we think that the counterfactual outcome would equal the observed outcome. That is, we impute $Y_i(1) = Y_i(0)$.

Step 3: Untreated cases (non-transitional truth commission cases)

We consider a unit treated only if it experiences a transitional truth commission, so units that experience non-transitional commissions are considered untreated. However, case study scholarship focusing on these non-transitional cases often provide clues about dynamics around truth and reconciliation options at the time of transition, allowing us to make guesses about what would have happened. In most scenarios where non-transitional truth commissions are set up many years later, we find that the demand for truth commissions existed even at the time of transition, but the reason for not establishing a truth commission was rarely a threat of renewed violence. Instead, the reason was often continued political infighting despite a formal transition, limited capacity amid other rebuilding concerns, leadership preferences, or conflict fatigue (USIP, 2011*b*; Vandeginste, 2012; Wiebelhaus-Brahm, 2009*a*). Such cases are imputed as $Y_i(0) = Y_i(1)$. Under these conditions, a truth commission would neither have made things worse nor made them better.

In some cases, however, a non-transitional truth commission comes up years after transition specifically because a TTC was rejected at the time of transition. In Uruguay, for instance, Roniger and Sznajder (1998) shows how a proposed truth commission was threatened by the military and rejected by the public at large. Specifically, he reports that "for years, the military opposed any opening of the issue of past human rights violations and did not acknowledge any institutional responsibility for abuses committed before 1985, threatening to destabilize the democratic situation if policies of accountability reached the state agenda." In addition, Uruguay had in place an impunity-imparting Law of Expiry that provided immunity to the members of security forces involved in acts that constituted violations of basic human rights. After a sustained human rights mass mobilization initiative at the domestic level, a referendum was conducted to decide whether to overturn the law. The results upheld the Law of Expiry by a margin of over 13 points (56.6% in favor and 43.3% against), with a turnout of 84.7%. While these results did not end public discussion and civil society pressure for accountability, these figures betray cognizance at the domestic level of the significant dangers that a truth commission could potentially pose if it were transitional in nature. As a result, this case is imputed as $Y_i(1) = 1$.

The imputation is strengthened by a comparison of this case to that of Brazil, which presents a natural match. In this context (also post-authoritarian like Uruguay), a similar Amnesty Law was in effect and was in fact credited with having opened the path to democratization in important ways (Schallenmueller, 2014). However, despite the law, Wiebelhaus-Brahm (2009*a*) shows that civilians did not reject the idea of a truth commission. Instead, they organized incremental demands for investigation of past human rights, though with very limited success because of the lack of military cooperation. In response, lawyers and other civilians sought the assistance of the Church along with victims' and human rights groups who documented human rights violations to collect information on cases brought before military courts. Based on this information, the São Paulo diocese published the investigation's final report, *Brasil: Nunca Mais*, a collection of allegations of over 1,800 cases of torture and murder committed since the military takeover in 1964. While this report was not officially endorsed nor did it cover all violations, it did not invite military backlash and allowed for human past rights abuses to remain in the public eye during the transition. These observations lead Wiebelhaus-Brahm (2009*a*) to conclude that in Brazil, "Groups frequently turn over information they have accumulated to truth commissions. A prospective Brazilian truth commission would likely be no different." This is not only a fitting representation of scholars of transitional justice making counterfactual guesses about what would have happened had there been a truth commission. It is also an indication that had an official truth commission been established, it would have been able to garner similar information without risking a return to military rule.

Step 4: Untreated cases

Lastly, we looked at cases that were truly untreated in that they never received a truth commission. This step represents the toughest case for imputation, given the lack of scholarship about truth commissions in scenarios where they never occured. As a result, this is the step from which the most cases are left unimputed. Within some of these untreated cases, the idea of truth commissions is sometimes brought up by civil society actors, opposition parties and international organizations but not acted upon. Some communities set up their own, unofficial truth-telling processes, while in others, we could find no evidence of discussion around transitional justice at all. Few studies actively address the reasons behind the failure of a truth commission to be set up, making our task more difficult.

A few examples clarify the type of evidence used to make imputations in this step. First, in post-conflict Bosnia, Dragovic-Soso (2016); Subotić (2010) find that multiple attempts to establish a truth commission failed for two main reasons. First, there was widespread resistance by conflict-era leaders to embrace social and political reconstruction. "Bosnian national leaderships expressed great willingness to support the TRC but only because they felt it was a good vehicle to tell their side of the story... [and] so they could kill it when it stopped serving their national interests" (Subotić, 2010, p. 147). Misplaced intentions, along with grievances such as inadequate inclusion in the discussion phase, in turn led domestic civil society organizations to unanimously reject the proposal for a truth commission. Second, institutional rivalry over primacy between the judicial International Criminal Tribunal for the former Yugoslavia (ICTY) and the truth commission project caused the former to view the commission as a competitor. Along with issues of the political viability of a truth commission in the face of limited international aid, "ICTY officials feared that the TRC could undermine the Tribunal's own investigations and decisions, witness protection programme and secret indictments" (Dragovic-Soso, 2016). In a counterfactual world then, an investigative truth commission would have likely been partial to state interests and would have probably clashed with the ICTY. However, such a commission would not have increased the potential for return to conflict given that accountability was being pursued by other means without backlash. As a result, the Bosnian case is coded as $Y_i(0) = Y_i(1) = 0$.

On the other hand, in many military-ruled contexts, the ousted military continues to wield disproportionate influence even after a formal transition is completed. In such a scenario, some experts have expressed concerns that a truth commission is exceedingly unlikely and if set up, could make fragile transitions more intractable. In Mynamar, for instance, the military junta not only perpetrated large-scale violence in the Arakan and Kachin conflicts, but also remained a powerful player despite the political reforms of 2011. Even after its transition to a civilian state, discussions around transitional justice have been met with well-documented fears of provoking backlash. For instance, in 2012, a group of state and civil society groups to discuss Burmese transitions argued that "Burma is not yet ready to follow in South Africa's footsteps by embarking on a path toward transitional justice...Such a move could even hinder the ongoing process of political reform in Burma" (Naing, June 15, 2012). Further, Holliday (2014) finds that key institutions are still struggling to find their feet, the judiciary remains deeply deficient and hence the pace of transitional justice in Myanmar is likely to be "sporadic, fragile and contested" (197). As a result, though some scholars have pitched truth commissions as the least threatening option Dukalskis (2015), most worry that pushing too hard on a transitional justice agenda could backfire. They conclude:"We can see nascent moves for transitional justice mechanisms as part of these reforms but no overall willingness or strategy to deal with the past ...there is a lot of ambivalence as well, and fear, that bringing up the past will provoke a coup by the military" (International Center for Transitional Justice, 2013). In light of these concerns, a transitional truth commission could not have been established under military rule (to investigate its own abuses), without risking a resumption to conflict. The treated potential outcome is therefore imputed as 1.

Finally, stakeholders from other contexts that did not establish truth commissions often assess potential TTCs by comparing their context to that of other successful commissions. In Angola, for instance, International Center for Transitional Justice (2008) finds that respondents often shared the view that although the TRC model was suited for the South African context, where the conflict's actors could easily be distinguished into perpetrators and victims, this was not the case in Angola where virtually the entire country was involved in the conflict. According to one of his respondents: 'Should we (Angolans) all sit at an Angolan TRC? How can we expect Angolans, exhausted from years and years of conflict, to even entertain such an idea?' The struggle of day-to-day existence leaves very little time for any other issues, including reconciliation. Moreover, concerns were raised that a TRC would taint the government's record as liberators, compromising the 'liberation discourse' so cherished by government since the end of the war." As discussed above, such analyses show that truth commissions are often not established for reasons unrelated to threats of return to conflict. In this case then, we followed the same logic and imputed $Y_i(1) = Y_i(0)$.

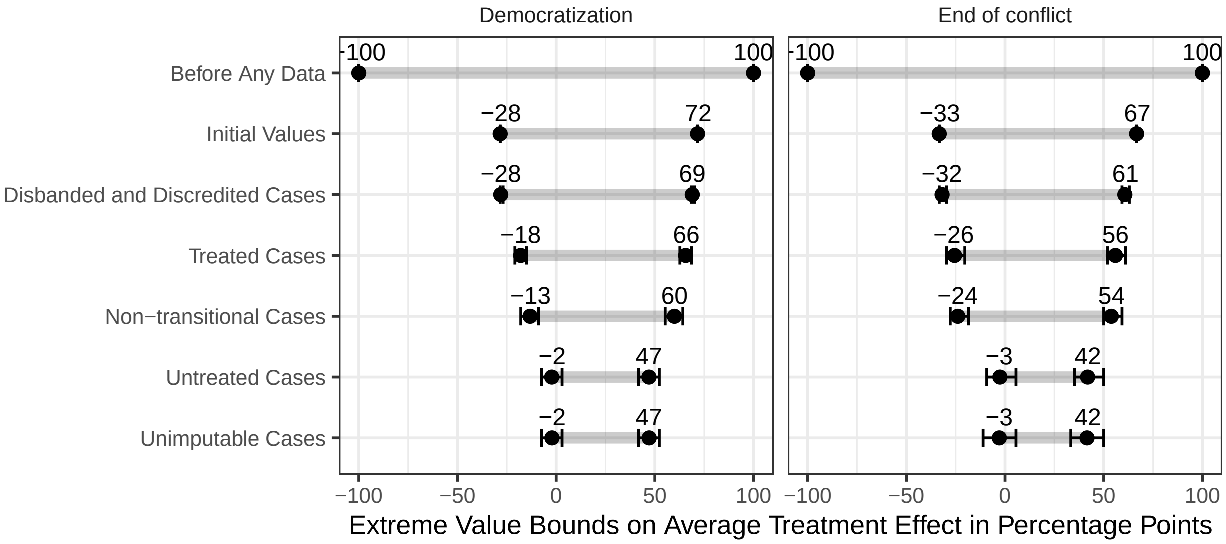
Summary

Figure 2 summarizes our results. Before any data collection, the extreme value bounds are 200 points wide, reflecting the total absence of information about the average effect of TTCs. After the observed data are collected, the width of the bounds shrink to 100 points because the world reveals half the potential outcomes. The four steps above shrink the uncertainty further as missing potential outcomes are filled in. For the end of conflict cases, the bounds come to [-4, 41] (only 45 points wide). In the democratization cases, the final bounds are 49 points wide.

For both sets of cases, the bounds include zero. The data and our state of knowledge are currently consistent with both positive, negative, and zero average effects. This is *very importantly* different from a "null" finding. The bounds are as wide as they are because we do not know as much about the effects of TTCs as we would like. The width of the bounds indicates either what work is left to be done or which counterfactuals are simply too unknowable to be imputed.

Table 3 reports the number of cases in which we think a TTC had or would have had a positive, negative, or zero effect, as well as the number of cases in which we were unable to make an imputation. In our view, the most important pattern is that in the majority of imputed cases, we find that the effect of truth commissions on our outcome variables is zero. Specifically, in

Figure 2: Transitional Truth Commissions: Extreme Value Bounds



41% of end of conflict cases and 43% of democratization cases, our imputations imply that truth commissions had or would have had no effect on the resumption of conflict or authoritarianism. Another important pattern is that we are unable to make imputations in about half the total number of cases, hence the bounds remain wide and the gaps in our knowledge persist. These bounds are representative of debates in the literature (Hirsch et al., 2012) about the average effects of truth commissions. Proponents find that truth commissions are essential to the consolidation of democracy and the resolution of conflict, a pragmatic middle road between prosecutions and impunity. They improve human rights protection, open up new avenues for activism and facilitate civil society participation (Kim and Sikkink, 2010; Bakiner, 2015; Wiebelhaus-Brahm, 2010; Borer, 2006). Critics, on the other hand, argue that the relationship between truth-telling and peace-building is flawed. Instead, truth commissions can have perverse effects by exacerbating tensions or providing smoke screens for abusive regimes (Mendeloff, 2004; Snyder and Vinjamuri, 2004). Given that our analysis is a summary of expert knowledge on the subject, it is appropriate that our extreme value bounds represent these ambiguities.

One of principal difficulties facing a traditional quantitative analysis of the effects of TTCs is that those units that come to be treated and untreated are so different from one another in both

Table 3: Summary of Imputations

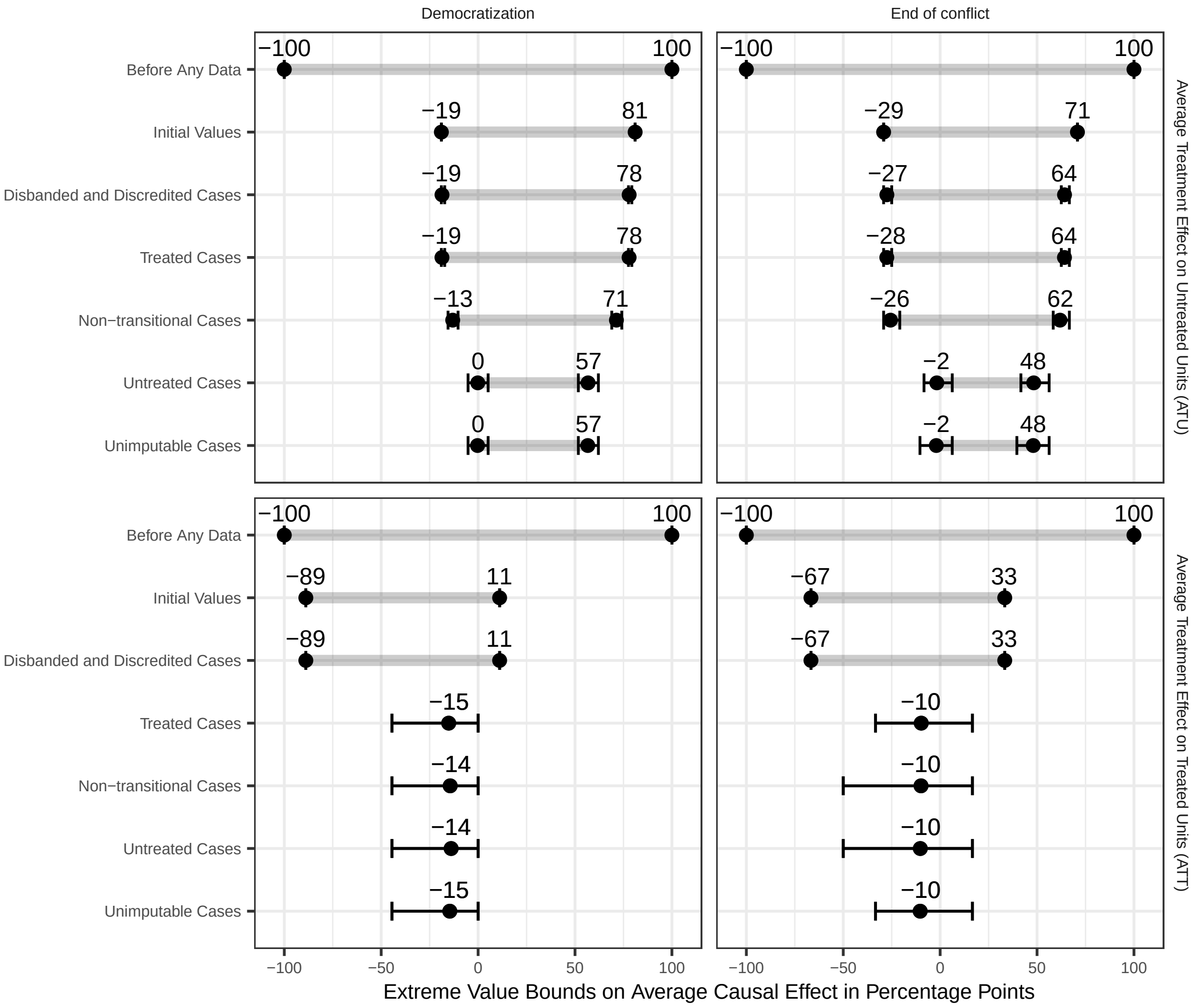
	Negative Effect	No Effect	Positive Effect	Unimputed
Democratization	1% (1)	43% (29)	6% (4)	49% (33)
End of conflict	2% (1)	41% (22)	13% (7)	44% (24)

observed and unobserved ways. Accordingly, the average treatment effect itself might not be the only interesting estimand – we might wish to know the average treatment effect among the treated (the ATT) or the average treatment effect among the untreated (the ATU, sometimes called the ATC where the C stands for "control"). The ATT and the ATU need not be equal, since units that select in or out of treatment may be quite different from one another.

Figure 3 shows how the extreme value bounds for the treated and untreated units develop differently. There are far more untreated units than treated units, and we know far less about them, again owing to the dearth of qualitative scholars digging into the cases that could have been treated with a TTC but were not. The bounds on the ATU are greater than 50 points wide, compared with the bounds on the ATT, which shrink all the way down to a point. We can summarize the ATT as a -11 percentage point effect on return to authoritarianism and a -14 percentage point effect on return to conflict. Further, even though the ATC is not bounded away from zero, it is bounded away from the ATT in the democratization cases. This analysis underlines that as a whole, units that do and do not come to be treated are different from one another, not just in their observable characteristics, but also in their responses to treatment.

We close this empirical example with two concluding thoughts. First, scholars of transitional justice are concerned not just with the effects of truth commissions on average, but also the conditions under which they are more or less effective. For instance, Olsen, Payne, Reiter and Wiebelhaus-Brahm (2010) find that truth commissions do not promote stability on their own but when used in combination with trials and amnesties, these commissions tend to have a positive impact. Our framework could be extended to consider such conditional hypotheses, by applying the method separately to cases that did and did not include such additional transitional justice tools. The resulting output would take the same form as Figure 3, where the conditioning variable is not the revealed treatment condition, but instead the presence or absence of, say, trials and amnesties.

Figure 3: Transitional Truth Commissions: Extreme Value Bounds



Second, we recognize that the resumption of conflict or return to authoritarianism are not the only dependent variables of interest. We chose these dependent variables because we consider them to be among the most important. In most cases that we are able to impute, we have found that TTCs have no effect on these outcomes, but that of course does not preclude effects on other outcomes we might have considered. For example, cross-national studies focusing on truth commissions have estimated the impact of TTCs on the level of democracy or repression, respect for or promotion of human rights and ability to inspire civil society action (Wiebelhaus-Brahm, 2010; Olsen, Payne, Reiter and Wiebelhaus-Brahm, 2010; Bakiner, 2015; Kim and Sikkink, 2010). As described below, our method could be straightforwardly extended to accommodate those (nonbinary) dependent variables as well.

Discussion

In this article, we have proposed a procedure that combines single-case qualitative inference with extreme value bounds. The main purpose of procedure is to summarize qualitatively-derived beliefs about average causal effects in a structured fashion. In cases where existing evidence is strong enough, we can impute counterfactual outcomes, which is equivalent to claiming knowledge of the individual treatment effect. In cases where existing evidence is weak, we can consider worst- and best-case scenarios in order to bound the ATE and to express our uncertainty about it.

We think this procedure will be most applicable to medium-N empirical questions about which quite a bit of previous knowledge has been generated. When the number of units is quite small, then a series of case studies is probably more appropriate than our method. When the number of units is much larger than a hundred, it becomes difficult to do justice to each particular case. This medium-N Goldilocks zone includes many topic areas in political science, including the study of leaders, states or nations, intra- or inter-state armed conflicts, treaties, or elections, to brainstorm a few.

We faced some rewarding difficulties applying our method to transitional truth commissions. One trouble was defining potential outcomes in the first place. For those units that experienced a TTC, it is clear what is meant by the "treated potential outcome." But what does it mean if those places did not experience a TTC? Would they have experienced lustrations or purges instead? There are many ways that each of these transitions could have evolved without TTCs and choosing just one for all treated cases was difficult. We followed Hernán and Robins (2016)'s notion of a "target trial," or what would have happened if at the moment of choosing to do a TTC, we had intervened to stop it. This notion helped us to specify the "closest possible world" as cleanly as possible, though this was not always easy.

Another difficulty was defining the universe of cases. We settled on including those cases that had, in our view, a probability of treatment that was not 0 or 1. This choice of course backs up the problem to deciding, among those cases that were not treated, whether they had a positive probability of being treated (and vice versa for the treated group). We acknowledge that others might make different choices about which cases were "eligible" for TTCs.

In the spirit of full transparency, we disclose that we realized halfway through data collection that the end of conflict and democratization cases could not be analyzed together because they did not share the same outcome variable. While we include this as a "difficulty" we faced, in fact, we felt liberated by the ability to split the analysis on qualitative grounds, rather than trying to force the two sets of units together, possibly by finding a dependent variable they shared in common that might have been less relevant for theory.

Of course, the main difficulty was making guesses about counterfactual states of the world. We were only able to engage with such a large set of cases because of the efforts of previous qualitative scholars of transitional justice. Our guesses about counterfactuals are summaries of our understanding of those works by others. We have laid out our reasoning for each case in the supplementary materials, but we are quite sure that others would dispute at least some of our guesses. We wholeheartedly welcome critical discussion of our imputations.

Stepping back from the application of transitional justice, we think there are many advantages to summarizing qualitative inferences in this way. First, we avoid conditioning our analyses on treated cases only. Because units are not randomly assigned to treatments, the average effect of treatment among the treated need not be the same as the average effect of treatment among the untreated. Our approach avoids the distortions associated with studying treated units only.

Second, we can explicitly account for the nonrandom selection into treatment. If units that do and do not receive treatment are different from each other in both observed and unobserved ways, comparing them (as in an observational, quantitative study) is inappropriate. A series of single-case qualitative studies considers each unit individually so worries about confounding are handled directly. Indeed, information about what make each unit special forms the basis for the qualitative inference.

Third, the process is transparent. If critics disagree about an imputation, they can offer a different one. If the disagreement is insoluble, we can simply remove the imputation altogether. Disputes over imputations underscore that we do not know the counterfactual, so it would be inappropriate to claim knowledge about a particular causal effect. In the worst case, we would have to leave all counterfactuals unimputed, which could only occur if qualitative case knowledge were entirely useless for causal inference. We are skeptical that we can compute counterfactuals in *all* cases, but we are optimistic that we can do so in at least *some* cases.

Finally, we think that this procedure can serve as a way of making disagreements among scholars explicit. Oftentimes, alternative readings of a case have so little in common that even locating the source of disagreement is difficult. Using this procedure requires that scholars state their best guess as to what would have happened to an outcome in particular. In this way, it encourages scholars to be bold in their causal claims.

References

Abadie, Alberto, Alexis Diamond and Jens Hainmueller. 2015. "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science* 59(2):495–510.

Acharya, Avidit, Matthew Blackwell and Maya Sen. 2016. "Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects." *American Political Science Review* 110(3):512–529.

Aronow, Peter M., Jonathon Baron and Lauren Pinson. 2017. "A Note on Dropping Experimental Subjects who Fail a Manipulation Check." *Political Analysis* . Forthcoming.

Bakiner, Onur. 2013. "Truth Commission Impact: An Assessment of How Commissions Influence Politics and Society." *International Journal of Transitional Justice* 8(1):6–30.

Bakiner, Onur. 2015. *Truth Commissions: Memory, Power, and Legitimacy.* University of Pennsylvania Press.

Beach, Derek and Rasmus Brun Pedersen. 2016. *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing.* University of Michigan Press.

Bennett, Andrew and Jeffrey T. Checkel, eds. 2014. *Process Tracing.* Cambridge: Cambridge University Press.

Blattman, Christopher and Edward Miguel. 2010. "Civil War." *Journal of Economic literature* 48(1):3–57.

Borer, Tristan Anne. 2006. *Telling the Truths: Truth Telling and Peace Building in Post-conflict Societies.* Univ of Notre Dame Pr.

Brady, Henry E. and David Collier. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Lanham, Maryland: Rowman & Littlefield Publishers.

Bullock, John G, Donald P Green and Shang E Ha. 2010. "Yes, But What's the Mechanism? (Don't Expect an Easy Answer)." *Journal of personality and social psychology* 98(4):550.

Collier, David. 2011. "Understanding Process Tracing." *PS: Political Science & Politics* 44(4):823–830.

Coppock, Alexander. 2018. "Avoiding Post-Treatment Bias in Audit Experiments." *Journal of Experimental Political Science* . Forthcoming.

Dancy, Geoff, Hunjoon Kim and Eric Wiebelhaus-Brahm. 2010. "The Turn to Truth: Trends in Truth Commission Experimentation." *Journal of Human Rights* 9(1):45–64.

Dragovic-Soso, Jasna. 2016. "History of a Failure: Attempts to Create a National Truth and Reconciliation Commission in Bosnia and Herzegovina, 1997–2006." *International Journal of Transitional Justice* 10(2):292–310.

Dukalskis, Alexander. 2015. "Transitional Justice in Burma/Myanmar: Crossnational Patterns and Domestic Context." *Irish Studies in International Affairs* 26:83–97.

Fairfield, Tasha. 2013. "Going Where the Money Is: Strategies for Taxing Economic Elites in Unequal Democracies." *World Development* 47:42–57.

Fearon, James D. 1991. "Counterfactuals and Hypothesis Testing in Political Science - Jstor." *World politics* 43(2):169–195.

Gary King, Robert O. Keohane, Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research.* Princeton University Press.

Geddes, Barbara, Joseph Wright and Erica Frantz. 2014. "Autocratic Breakdown and Regime Transitions: A New Data Set." *Perspectives on Politics* 12(2):313–331.

George, Alexander L and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences.* mit Press.

Gerber, Alan S. and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation.* New York: W.W. Norton.

Gerring, John. 2006. *Case Study Research: Principles and Practices.* Cambridge university press.

Gerring, John. 2010. “Causal Mechanisms: Yes, But?” *Comparative political studies* 43(11):1499–1526.

Gibson, James L. 2002. “Truth, Justice, and Reconciliation: Judging the Fairness of Amnesty in South Africa.” *American Journal of Political Science* pp. 540–556.

Gibson, James L. 2004. “Does Truth Lead to Reconciliation? Testing the Causal Assumptions of the South African Truth and Reconciliation Process.” *American Journal of Political Science* 48(2):201–217.

Gibson, James L. 2006. “Overcoming Apartheid: Can Truth Reconcile a Divided Nation?” *The Annals of the American Academy of Political and Social Science* 603(1):82–110.

Glynn, Adam N. and Nahomi Ichino. 2015. “Using Qualitative Information to Improve Causal Inference.” *American Journal of Political Science* 59(4):1055–1071.

Goertz, Gary and James Mahoney. 2012. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton University Press.

Haber, Stephen and Victor Menaldo. 2011. “Do natural resources fuel authoritarianism? A reappraisal of the resource curse.” *American political science Review* 105(1):1–26.

Harvey, Frank P. 2012. “President Al Gore and the 2003 Iraq War: A Counterfactual Test of Conventional ‘W’isdom.” *Canadian Journal of Political Science/Revue canadienne de science politique* 45(1):1–32.

Hayner, Priscilla B. 2000. *Unspeakable Truths: Confronting State Terror and Atrocity*. Routledge.

Hayner, Priscilla B. 2006. “Truth Commissions: A Schematic Overview.” *International Review of the Red Cross* 88(862):295–310.

Hernán, Miguel A and James M Robins. 2016. “Using Big Data to Emulate a Target Trial When a Randomized Trial is Not Available.” *American journal of epidemiology* 183(8):758–764.

Hirsch, Michal Ben-Josef, Megan MacKenzie and Mohamed Sesay. 2012. "Measuring the Impacts of Truth and Reconciliation Commissions: Placing the Global 'Success' of TRCs in Local Perspective." *Cooperation and Conflict* 47(3):386–403.

Höglund, Kristine and Camilla Orjuela. 2011. "Winning the Peace: Conflict Prevention After a Victor's Peace in Sri Lanka." *Contemporary Social Science* 6(1):19–37.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396):945–960.

Holliday, Ian. 2014. "Thinking About Transitional Justice in Myanmar." *South East Asia Research* 22(2):183–200.

Hume, David. 1748. *An Enquiry Concerning Human Understanding.* Oxford: Oxford University Press.

Humphreys, Macartan and Alan M. Jacobs. 2015. "Mixing Methods: A Bayesian Approach." *American Political Science Review* 109(4):653–673.

Imai, Kosuke, Luke Keele, Dustin Tingley and 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.

International Center for Transitional Justice. 2008. "Southern African Regional Assessment Mission Report: Angola.".

International Center for Transitional Justice. 2013. "ICTJ Program Report: Myanmar.".
URL: *https://www.ictj.org/news/ictj-program-report-burma-myanmar*

International Crisis Group. 22 December 2011. "Statement on the Report of Sri Lanka's Lessons Learnt and Reconciliation Commission.".

Kim, Hunjoon and Kathryn Sikkink. 2010. "Explaining the Deterrence Effect of Human Rights Prosecutions for Transitional Countries." *International Studies Quarterly* 54(4):939–963.

Kreutz, Joakim. 2010. "How and When Armed Conflicts End: Introducing the UCDP Conflict Termination dataset." *Journal of Peace Research* 47(2):243–250.

Laplante, Lisa J and Kimberly Theidon. 2007. "Truth with Consequences: Justice and Reparations in Post-Truth Commission Peru." *Human Rights Quarterly* 29:228.

Lebow, Richard Ned. 2010. *Forbidden Fruit: Counterfactuals and International Relations.* Princeton University Press.

Lebow, Richard Ned and Janice Gross Stein. 1996. "Back to the Past: Counterfactuals and the Cuban Missile Crisis." *Tetlock and Belkin (Eds.)* pp. 119–148.

Lewis, David. 1973. Counterfactuals and comparative possibility. In *Ifs.* Springer pp. 57–85.

Lewis, David. 1979. "Counterfactual Dependence and Time's Arrow." *Noûs* pp. 455–476.

Lonergan, Kate. 2017. "Does Reconciliation Prevent Future Atrocities." *Evaluating Pratice in Sri Lanka. United States Institute for Peace (USIP): Peaceworks* (132).

Mahoney, James. 2010. "After KKV: The New Methodology of Qualitative Research." *World Politics* 62(1):120–147.

Mahoney, James. 2012. "The Logic of Process Tracing Tests in the Social Sciences." *Sociological Methods and Research* 41(4):570–597.

Manski, Charles F. 1999. *Identification Problems in the Social Sciences.* Harvard University Press.

Mendeloff, David. 2004. "Truth-Seeking, Truth-Telling, and Postconflict Peacebuilding: Curb the Enthusiasm?" *International Studies Review* 6(3):355–380.

Morgan, Stephen L and Christopher Winship. 2014. *Counterfactuals and Causal Inference.* Cambridge University Press.

Naing, Saw Yan. June 15, 2012. "Burma 'Not Ready' for Truth Commission." *The Irrawadi* .

Neyman, Jerzy. 1923. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. Reprint, 1990." *Statistical Science* 5(4):465–472. with Dabrowska, Dorota M. and Speed, Terence P.

Nwogu, Nneoma V. 2007. *Shaping Truth, Reshaping Justice: Sectarian Politics and the Nigerian Truth Commission.* Lexington Books.

Olsen, Tricia D, Leigh A Payne and Andrew G Reiter. 2010. "Transitional Justice in the World, 1970-2007: Insights From a New Dataset." *Journal of Peace Research* 47(6):803–809.

Olsen, Tricia D, Leigh A Payne, Andrew G Reiter and Eric Wiebelhaus-Brahm. 2010. "When Truth Commissions Improve Human Rights." *International Journal of Transitional Justice* 4(3):457–476.

Pearl, Judea. 2009. *Causality.* Cambridge university press.

Ragin, Charles C. 2014. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies.* Univ of California Press.

Rohlfing, Ingo. 2012. *Case Studies and Causal Inference: An Integrative Framework.* Palgrave Macmillan.

Roniger, Luis and Mario Sznajder. 1998. "The Politics of Memory and Oblivion in Redemocratized Argentina and Uruguay." *History and Memory* 10(1):133–169.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66(5):688.

Schallenmueller, Christian Jecov. 2014. Transitional Justice in Brazil and Uruguay: Different Solutions to the Tension Between Human Rights and Democracy. In *Glasgow, European Consortium for Political Research General Conference.* Vol. 2014 pp. 10–11.

Seawright, Jason. 2016. *Multi-Method Social Science: Combining Qualitative and Quantitative Tools.* Cambridge University Press.

Skaar, Elin. 1999. “Truth Commissions, Trials-or Nothing? Policy Options in Democratic Transitions.” *Third world quarterly* 20(6):1109–1128.

Snyder, Jack and Leslie Vinjamuri. 2004. “Trials and Errors: Principle and Pragmatism in Strategies of International Justice.” *International security* 28(3):5–44.

Subotić, Jelena. 2010. *Hijacked Justice: Dealing With the Past in the Balkans.* Cornell University Press.

Tetlock, Philip E and Aaron Belkin. 1996. *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives.* Princeton University Press.

Thiranagama, Sharika. 2013. “Claiming the State: Postwar Reconciliation in Sri Lanka.” *Humanity: an International Journal of Human Rights, Humanitarianism, and Development* 4(1):93–116.

USIP. 2011*a*. “Truth Commission: Bolivia - United States Institute of Peace.”.
URL: *https://www.usip.org/publications/1982/10/truth-commission-bolivia*

USIP. 2011*b*. “Truth Commission Digital Collection - United States Institute of Peace.”.
URL: *https://www.usip.org/publications/2011/03/truth-commission-digital-collection*

Vandeginste, Stef. 2012. “Burundi’s Truth and Reconciliation Commission: How to Shed Light on the Past While Standing in the Dark Shadow of Politics?” *International Journal of Transitional Justice* 6(2):355–365.

Wiebelhaus-Brahm, Eric. 2009*a*. “What Does Brazil Have to Gain from a Truth Commission after Two Decades of Democracy?” *International Conference on the Right to Truth, Sao Paulo, Brazil.* .

Wiebelhaus-Brahm, Eric. 2009*b*. “What is a Truth Commission and Why Does it Matter?” *Peace and Conflict Review* 3(2):1–14.

Wiebelhaus-Brahm, Eric. 2010. *Truth Commissions and Transitional Societies: The Impact on Human Rights and Democracy.* Routledge.

Woodward, James. 2005. *Making Things Happen: A Theory of Causal Explanation.* Oxford university press.

Yasmin Sooka, Frances Harrison. 2019. "Why Has Sri Lanka's Transitional Justice Process Failed to Deliver?".
URL: *https://blogs.lse.ac.uk/southasia/2019/02/06/long-read-why-has-sri-lankas-transitional-justice-process-failed-to-deliver/*

Yusuf, Hakeem O. 2007. "Travails of Truth: Achieving Justice for Victims of Impunity in Nigeria." *The International Journal of Transitional Justice* 1(2):268–286.