Civil resistance in the streetlight:
Replicating and assessing evidence on nonviolent effectiveness

Christoph Dworschak

University of York
Norwegian University of Science and Technology

christoph.dworschak@york.ac.uk

5th January 2023

Does civil resistance work? Research emphasizes the effectiveness of nonviolent resistance over violent resistance in achieving campaign goals, with the seminal study "Why Civil Resistance Works" (WCRW) by Chenoweth and Stephan being the main point of reference to date. I revisit this pivotal finding in three steps. First, I reproduce WCRW's results on nonviolent effectiveness. Second, I discuss how cases may have been overlooked due to a streetlight effect. Third, I quantify the results' sensitivity using simulations. I find that WCRW's main findings on nonviolent effectiveness are highly sensitive to variable selection and undercoverage bias, bootstrapping, and omitted variable bias. As a routine reference in scholarship and the public discourse, assessing the robustness of WCRW's findings is relevant to practitioners and research spanning the past decade.

Keywords: civil resistance, campaign success, violence bias, simulation.

I am grateful for the reports of two anonymous reviewers that helped to substantially improve this manuscript. I thank Charles Butcher, Scott Gates, Kristian Skrede Gleditsch, Amélie Godefroidt, Thea Johansen, Monika Onken, Christian Oswald, Jonathan Pinckney, Christoph V. Steinert, and the participants at the annual meeting of the Conflict Research Society, the Jan Tinbergen European Peace Science Conference, the Norwegian National Political Science Conference, the Securing the Victory workshop at the Peace Research Institute Oslo, the Workshop on Quantitative Political Research at the University of York, and the peace research seminar at NTNU for many helpful conversations and suggestions. I am indebted to Erica Chenoweth for their generous feedback and support. All mistakes are my own.

1 Introduction

Seminal research on civil resistance[1] finds that nonviolent tactics significantly outperform violent tactics in generating campaign success (Chenoweth and Stephan 2011; Stephan and Chenoweth 2008). Research of the past decade builds on this finding, with recent studies showing nonviolent resistance campaigns to have better prospects for swaying public opinion, and for generating long-term democratization, compared to violent campaigns (e.g., Pinckney 2020; Lambach et al. 2020; Orazani and Leidner 2019; Kim and Kroeger 2019). Meanwhile, an increasing number of studies add important qualification, identifying contexts in which nonviolent tactics may not be more effective in reaching campaign goals (Manekin and Mitts 2022; Gleditsch et al. 2021; Pischedda 2020; Chenoweth 2020). Instead of exploring context-specific limitations to the effectiveness of nonviolent resistance, I propose a re-examination of the core evidence for the notion of nonviolent success as a whole.

First, I reproduce the core quantitative findings on the success of nonviolent resistance in the pioneering book "Why Civil Resistance Works" (hereafter WCRW) by Chenoweth and Stephan, which constitutes the main reference for researchers, activists, and journalists arguing for nonviolent effectiveness to date. Second, I discuss the potential for campaigns having been overlooked in the data underlying WCRW, and how this may bias existing conclusions about the effectiveness of civil resistance. Specifically, among all those resistance campaigns that are unsuccessful, violent and casualty-intensive campaigns are more likely to receive international attention and get recorded than those that are nonviolent. This streetlight effect[2] might imply a systematic undercount of unsuccessful nonviolent campaigns. Third, I quantify how sensitive the results are to such missingness through simulations. I find that the book's main findings may be overturned, on average, by adding one nonviolent unsuccessful campaign to the original analysis sample. This results, as well as additional checks included in the Online appendix, suggest that the existing comparative evidence for nonviolent effectiveness provided in WCRW is not conclusive. Specifically, the findings lack statistical significance, and are not robust to bootstrapping, sample selection bias, and omitted variable bias. I conclude with a discussion of the implications for the study of civil resistance effectiveness.

This research note is not the first to raise the issue of sample selection in WCRW. Chenoweth and Stephan provide a useful discussion of the problem themselves and took important steps to mitigate it, and subsequent research follows up by highlighting this and related concerns (Anisin 2021; Onken, Shemia-Goeke and Martin 2021; Anisin 2020; Chenoweth, Pinckney and Lewis

1. Civil resistance is used synonymously with "nonviolent resistance" and "nonviolence" throughout this manuscript. All these instances refer to civil resistance as a method of contention as it is defined and used in Chenoweth and Stephan (2011), and not to "nonviolence" as the broader philosophical concept (see e.g., Chenoweth (2016) for a discussion).

2. The name is based on a short story about a policeman and a drunk person, which may be best summarized in short by the exchange: "Are you sure you lost your keys here?" – "No, officer, I lost them over there, but the light is much better here."

2018; Chenoweth 2016; Day, Pinckney and Chenoweth 2015; Lehoucq 2015; Chenoweth and Cunningham 2013). However, despite a general understanding that missingness is likely, it is unknown whether and how it may affect existing findings. A focused engagement with WCRW allows me to theorize and to quantify the degree to which systematic missingness may infringe on its influential findings. In doing so, I also uncover other sources of uncertainty, including variable selection, sample idiosyncrasies, and omitted variable bias.

Re-examining core evidence on civil resistance success by replicating and dissecting WCRW is an important contribution to the fields of Comparative Politics and International Relations. Winner of the American Political Science Association's prestigious Woodrow Wilson Prize, WCRW is one of the most influential works in peace science and fundamentally shaped the comparative research agenda on civil resistance. Ten years on, WCRW continues to serve as chief reference on nonviolent effectiveness for academics and laypeople alike, as it provides the most comprehensive comparative assessment of the success of nonviolent resistance to date. Other research on civil resistance effectiveness either focuses on more specific outcomes, like democratization, or is of more limited generalizability and inferential scope. Therefore, revisiting the established notion of civil resistance effectiveness needs to start by engaging WCRW's findings. Meanwhile, its pivotal role makes re-examining key elements of WCRW a reason on its own, contributing to social sciences' paramount quest for improving research transparency through replication.

2 The limelight: Replicating the success of civil resistance

WCRW develops a comprehensive theoretical framework linking civil resistance to campaign success (Chenoweth and Stephan 2011, ch. 2). Civil resistance is argued to attract higher levels of participation, because nonviolent action lowers physical, informational, moral, and commitment barriers compared to violent action. This ability of civil resistance to attract a larger and more diverse support base increases its opportunities to drain the regime's capacity. Notwithstanding potential counterarguments that are discussed along the way, WCRW concludes through a combination of theoretical considerations and case examples, backed up by descriptive and inferential statistics, that nonviolent action is more likely to generate campaign success than violent action. The key empirical implication of this theoretical framework is that nonviolent resistance movements are more effective in reaching their campaign goals than violent resistance movements.

WCRW tests this implication using multivariate regression analyses and finds support for civil resistance effectiveness.[3] The analyses are based on a dataset capturing the traits of 323

3. Chenoweth and Stephan (2011) present a host of quantitative and qualitative evidence in support of their argument. Among the quantitative tests I focus on the multivariate regression analyses, because they constitute the most comprehensive set of evidence. For a discussion of some of the other quantitative tests and of the scope conditions relating to the use of the book's original NAVCO 1.1 data, see the appendix section A.2.

campaigns between 1900 and 2006. It was published as NAVCO 1.1 as part of the Nonviolent and Violent Campaigns and Outcomes (NAVCO) data project. The unit of analysis is the individual campaign. The outcome of interest is whether a campaign succeeded, or failed, in reaching its policy objectives. Policy objectives usually refer to maximalist claims, like anti-regime, anti-occupation, and secessionist campaigns (except in the case of 11 campaigns, for which no such goal is defined). Similar to the outcome variable, the explanatory variable of interest is a dichotomous (non-)violence indicator (*nonviolent campaign*). It captures whether a campaign used predominantly nonviolent methods in the context of civil disobedience, or whether it employed violence in the context of an insurgency or revolution. In its main analyses, WCRW draws on a number of covariates to mitigate confounding and isolate the effect of civil resistance on campaign success.

In the following, I reproduce all eight regression models reported in Chenoweth and Stephan (2011, 70-71, Table 3.1). Replication is of fundamental importance to the reliability of empirical political science (King 1995) – an importance that has gained notoriety in the wake of the continuing replication crisis, highlighting the pivotal role of reproducibility as part of the scientific method. Leading by example, Chenoweth and Stephan (2011) provide the full replication data and code necessary for reproducing the analyses conducted in WCRW, which I draw on below. The results from these multivariate regression analyses, in combination with several insightful qualitative case studies, form the basis for the influential claim that nonviolent resistance is more successful than violent resistance. The models test the effect of *nonviolent campaign* on the likelihood of success under varying model specifications. Showing multiple models with different specifications enables the reader to better understand the robustness of the effect of *nonviolent campaign*, and whether its statistical significance is dependent on any specific modeling choice. The last model, Model 8, shows the effect of *nonviolent campaign* under the most extensive specification, conditioning on all covariates mentioned above. All other models are nested versions of Model 8. Therefore, I focus my discussion and later exploration on Model 8 and Model 7, with the latter being similarly comprehensive albeit excluding fixed effects.

The replication results are shown in Table 1. To facilitate comparison, the first row shows the replication goal, i.e., the original results as they appear printed in WCRW (ibid., 70-71). The second row shows the results generated based on the replication material accompanying Chenoweth and Stephan (2011).[4] The coefficients and standard errors are fully reproducible.[5] The difference in significance levels indicated for the main variable of interest, *nonviolent campaign*, in Models 2, 3,

4. WCRW replication data and script were retrieved from Erica Chenoweth's website at ericachenoweth.com/research on 11.07.2021. While the replication code is provided for STATA, I use R to facilitate the display of the results and later simulations, as well as to increase open-source accessibility. I show the full replication results of the main models 7 and 8 using both STATA and R in Table A.1 on page A-1 in the appendix, which demonstrate equivalence across platforms.

5. There are few minor deviations in some coefficients and standard errors among the covariates due to rounding, as well as a typo related to the decimal degree of the intercept in Model 7. These deviations are of no further relevance and are only mentioned here for formal completion of the replication.

Table 1: Replication of WCRW Table 3.1

	Dependent variable:							
	Campaign success							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Original:	0.90*	0.52***	0.43***	1.08***	1.26***	1.08***	0.96*	0.43***
Nonviolent camp.	(0.49)	(0.43)	(0.43)	(0.25)	(0.26)	(0.28)	(0.53)	(0.68)
Replication:	0.91*	0.52	0.43	1.09***	1.26***	1.08***	0.96*	0.43
Nonviolent camp.	(0.49)	(0.43)	(0.43)	(0.25)	(0.26)	(0.28)	(0.53)	(0.68)
Target polity score	✓						✓	✓
# participants, log	✓	✓	✓				✓	✓
Population, log	✓	✓	✓				✓	✓
Target capabilities		✓					✓	✓
Violent reg. repress.			✓				✓	✓
Secessionist camp.				✓			✓	✓
Anti-occupat. camp.				✓			✓	✓
Reg. change camp.				✓			✓	✓
Continent FEs					✓			✓
Decade FEs						✓		✓
Observations	141	153	163	323	323	323	134	134
Cluster-rob. SEs	✓	✓	✓	✓	✓	✓	✓	✓

Note: *p<0.1; **p<0.05; ***p<0.01

The first row, "Original", shows the coefficients, standard errors, and significance levels of *nonviolent campaign* as they appear printed in WCRW Table 3.1, page 70. The second row, "Replication", shows the coefficients, standard errors, and significance levels of *nonviolent campaign* as they are produced based on WCRW's replication materials.

and 8 is explained in WCRW in an accompanying footnote (p. 71) as "Nonviolent resistance and logged participants are jointly significant at p = .000." This suggests that a high significance level is printed next to the insignificant coefficient estimates of *nonviolent campaign* due to the significance of the linear combination of both *nonviolent campaign* and *number of participants*.[6] The same linear combination was not used to indicate significance levels next to the significant coefficient estimates of *nonviolent campaign* in Models 1 and 7.

The lack of significance in some of the models signals that the effect of nonviolence on campaign success is not robust to different model specifications,[7] and suggests that there may be a relevant degree of uncertainty over whether choosing nonviolent means over violent ones improves the prospect for campaigns' success. This does not mean that civil resistance does *not* work, but that findings based on the analyses in WCRW are inconclusive. Importantly, the results obtained

6. While this linear combination does not provide the total or direct effect of *nonviolent campaign*, nor its indirect effect through *number of participants*, it signals an intention to interpret *number of participants* as part of the effect of *nonviolent campaign*. I explore the role of *number of participants* in more detail in appendix section A.5, the results of which are summarized further below.

7. In the original analyses, differences due to variable selection cannot be discerned from differences due to varying analysis samples. Holding the sample constant (equal list-wise deletion across all eight models), however, the differences in statistical significance remain the same.

here are mere reproductions of the original models and data, without yet subjecting the findings to any additional stress tests.

3 The streetlight: Examining the role of missingness

The above analyses constitute the main comparative evidence for a broadly conceived notion of civil resistance effectiveness to date. Does this evidence bear scrutiny? Chenoweth and Stephan (2011) and subsequent research identify important limitations to the NAVCO 1.1 dataset, which forms the empirical basis for the above findings. These include a limited temporal coverage (Anisin 2020), difficulties in conceptualizing (non-)violence as primary resistance method (Onken, Shemia-Goeke and Martin 2021; Anisin 2020; Lehoucq 2015; Day, Pinckney and Chenoweth 2015), and relevant campaigns that were overlooked and thus not included in the data (Anisin 2020; Lehoucq 2015; Day, Pinckney and Chenoweth 2015). Despite the scholarly awareness of these limitations, there is no systematic assessment of the robustness of WCRW's finding on civil resistance effectiveness.

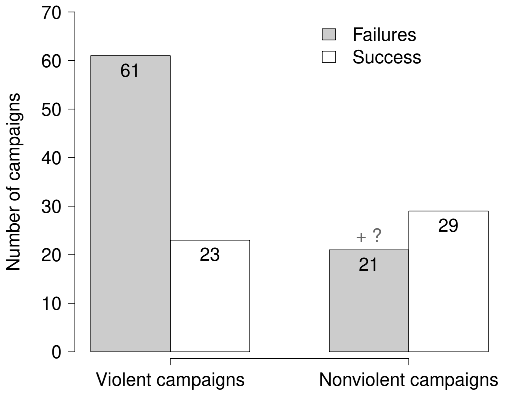
In light of the paramount relevance of WCRW's findings for both scholars and practitioners, these critiques merit systematic inspection. A single study cannot assess all potential issue areas, so I specifically focus on the likely omission of relevant campaigns: while it is known that some campaigns have been overlooked in the WCRW data (cf. Anisin 2020; Chenoweth 2016; Lehoucq 2015), it is unknown how sensitive WCRW's main findings are to the inclusion of these missing cases. Therefore, I begin by discussing the sampling frame of WCRW and how it might introduce missingness. I highlight why especially unsuccessful nonviolent campaigns may have been overlooked in the NAVCO v1 data series due to a streetlight effect. This is followed by analyses that quantify the robustness of the main findings in WCRW to such missingness.

Violence in the streetlight

Observational research cannot manipulate the treatment and, instead, has to make do with the data that is available. This makes observational research prone to various sampling biases, including the streetlight effect (undercoverage bias): data availability determines which observations make it into the sample, which can lead to the exclusion of relevant cases. A common example are surveys that miss respondents from hard-to-reach populations. If this missingness is systematic, it can distort or nullify true population patterns, or it can produce spurious patterns that do not exist in the underlying population.

When it comes to historical data on violent and nonviolent campaigns, it is usually the violence that is in the streetlight, with coverage of nonviolent resistance lagging behind the coverage of violent activity (Chenoweth and Cunningham 2013). Violence is easier to observe, is more readily recorded and tracked, and tends to receive more attention than civil resistance (Earl et al. 2004). This is also labeled "violence bias" (Day, Pinckney and Chenoweth 2015). For the purpose of analyzing the effect of civil resistance on campaign success, such an underrepresentation of nonviolent campaigns would be of little concern if the unobserved nonviolent campaigns were uniformly distributed across relevant strata. However, campaigns that were successful in achieving political transformation are likely recorded independent of their methods. Whether violent or nonviolent, a campaign that achieves maximalist policy change usually "makes the news" and enters the annals of history, and thus enters the dataset underlying WCRW.[8] Consequently, data on nonviolent campaigns is likely skewed due to a "success bias": civil resistance activity with maximalist goals gets recorded when it reaches a certain degree of maturity, but escapes international attention more easily if it is not able to reach that point (Chenoweth, Pinckney and Lewis 2018, 525). Figure 1 provides a visual summary of the empirical distribution of success and failure over violent and nonviolent resistance methods in WCRW's analysis sample, with the question mark indicating the empirical implication of the theorized omission of nonviolent failures.

Figure 1: WCRW sample of campaigns



Empirical distribution of campaign success and resistance methods in WCRW's Models 7 and 8, Table 3.1. The gray questionmark indicates the concern for potential missingness of nonviolent failures.

Addressing the question of missingness in WCRW's sample based on qualitative case knowledge, Lehoucq (2015) argues that several South American campaigns were overlooked that would have fulfilled NAVCO inclusion parameters. Of the 13 missing cases in South America, Lehoucq (2015) finds eleven to be nonviolent, in line with the suspected "violence bias" (Day, Pinckney and Chenoweth 2015). However, importantly, of these eleven nonviolent campaigns only three are listed as successful. In a response to Lehoucq (2015), Chenoweth (2016) acknowledges the omission of some cases from the WCRW sample, as is common for first-version datasets, but also notes that inclusion criteria in Lehoucq (2015) may have been different from the ones applied in WCRW. Following up on these discussions, Anisin (2020, p. 1127, emphasis in original) summarizes the issue as "While few data sets or projects result in the collection of a true universe of cases, the issue with [the NAVCO data project] is that it is *adversely incomplete* because a number of failed nonviolent revolutions are left out as are a plethora of successful violent revolutions."[9] While it may be

8. See appendix section A.3 for an extended discussion of the source materials used in WCRW.

9. While this goes beyond the scope of this study, a higher rate of missing successful violent revolutions would further exacerbate the undercoverage bias.

impossible to determine the exact number and distribution of omitted cases, these previous studies provide a qualitative baseline to motivate further inquiry. Specifically, although they convey a general concern that WCRW's sample may have omitted unsuccessful nonviolent campaigns, it remains to be examined whether an inclusion of these overlooked cases in WCRW's analyses would infringe upon its main findings.

Chenoweth and Stephan (2011, 79, and web appendix, 8-9) offer a detailed discussion of the potential for nonviolent campaigns that did not mature enough to get recorded. They took several steps towards mitigating this risk, and conclude by simulating an extreme scenario in which half of the violent campaigns pose as failed nonviolent campaigns. Shifting the base rates in such a way, Chenoweth and Stephan (2011, 79) discuss how even this change is not enough to make violence more effective than nonviolence in this bivariate setup. While this is a useful first step towards stress testing their results, it assumes a bivariate setup and does not account for the likely confounding influences discussed throughout their study. Therefore, below I present a systematic examination of the role of missingness for WCRW's main findings.

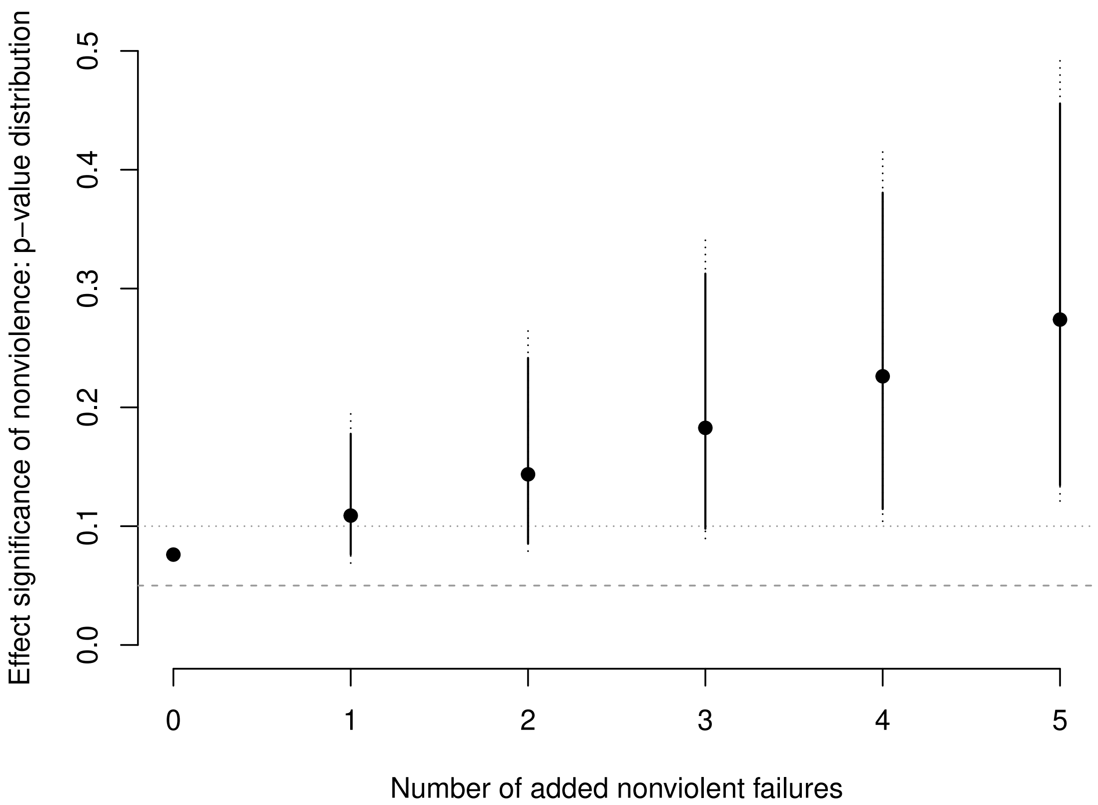
Simulating the effect of missingness

If there were failed nonviolent campaigns that were overlooked and not recorded, would the finding on civil resistance effectiveness change if we included them in the analysis? I quantify the results' sensitivity to unobserved nonviolent failures by simulating failed nonviolent campaigns and adding them to the original data. This way, I probe how many overlooked campaigns it would take to render the effect of nonviolence statistically indistinguishable from zero. I use Model 7 (cf. Table 1 on page 5) as basis for this stress tests. Model 7 is the most comprehensive model that is still statistically significant, with the other models being nested versions of it. Running stress tests on the nested models is less meaningful, as they do not partial out rivaling explanations to the same extent. Probing the sensitivity of Model 8 is not as useful either, since its results are not statistically significant to begin with.

To recreate the hypothesized data generating process of the missingness as close as possible, I simulate additional failed nonviolent campaigns and add them to the analysis data. In appendix section A.4, I also provide an approach in which I bootstrap the original analysis sample while incrementally adjusting the sampling weights to over-sample failed nonviolent campaigns.[10] To ensure that these simulated failed campaigns are representative of real campaigns, I make use of the remaining WCRW data that were list-wise deleted from the analysis sample due to missing information in some of the covariates. These omitted data include 28 failed nonviolent campaigns. I pad the

10. I find that a standard bootstrapping of the sample of Model 7 turns the effect of nonviolence statistically insignificant even before making any further adjustments to the underlying sampling weights.

Figure 2: Results adjustment assuming unobserved nonviolent failures



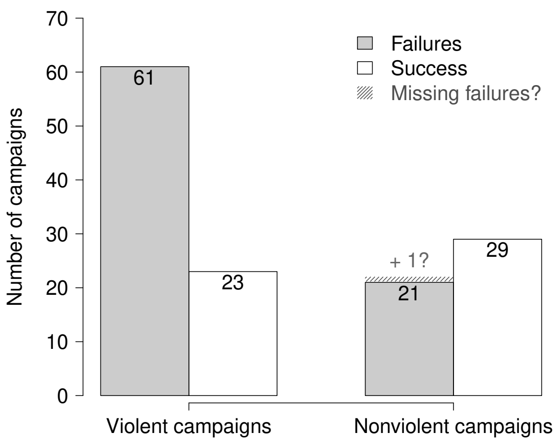
Simulated *p*-value distributions of the effect of *nonviolent campaign*. Simulations are based on adding existing, list-wise deleted observations to the analysis sample. The data and model specification are based on Model 7 in Table 1. Added nonviolent failures are drawn randomly 10,000 times with replacement from a pool of 1400 permutations. Points indicate distribution means, solid lines the [0.05, 0.95] quantile intervals, and dotted lines the [0.025, 0.975] quantile intervals. Horizontal lines indicate the 0.05 and 0.1 significance thresholds respectively. The left-most entry shows the original *p*-value of Model 7 (n = 134). All uncertainty estimates are based on cluster-robust standard errors, in line with the original model.

missing cells using multiple imputation (Honaker, King and Blackwell 2011; King et al. 2001), simulating 50 different datasets of the 28 nonviolent failures.[11] These 1400 observations serve as pool from which the additional failed nonviolent campaigns are sampled, generating many slightly adjusted datasets. For example, when simulating that two failed nonviolent campaigns were overlooked, the original analysis sample is combined with two random campaigns from the 1400 nonviolent failures. This is repeated 10,000 times. Each time, Model 7 is run on this padded sample, giving as many possible analysis results.

11. As a robustness check, instead of imputing missing covariate values of list-wise deleted observations, I draw on updated versions of the book's original data sources (Feenstra, Inklaar and Timmer 2015; Singer, Bremer and Stuckey 1972) to fill in the missing information. The results do not substantially differ from those visualized in Figure 2 and are visualized in the Online Appendix, Figure A.1 on page A-2.

The results of this sensitivity analysis are visualized in Figure 2. The left-most point shows the effect significance[12] of *nonviolent campaign* of the original Model 7 as it appears in the replication above (n = 134). Each step to the right adds one hypothetical failed nonviolent campaign to the analysis sample (up to n = 139). These are distributions of *p*-values, with the points representing the means, the solid lines the [0.05, 0.95] quantile intervals, and the dotted lines the [0.025, 0.975] quantile intervals. The dispersion is due to uncertainty, which enters the analyses in two ways: first, the multiple imputation generates a range of 50 likely covariate values for each missing cell, all of which result in slightly different effect estimates. Second, the bootstrap sampling leads to a different set of hypothetical failed nonviolent campaigns being added to the original analysis sample in each iteration. Therefore, the points and intervals give an indication of the most likely effect significance, as well as the spread of its less likely realizations. With the majority of simulated *p*-values surpassing the 0.1 threshold, the results suggest that just one overlooked failed nonviolent campaign may suffice for the effect of *nonviolent campaign* to not be statistically distinguishable from zero anymore (Figure 3).

Figure 3: Sensitivity of existing evidence



Based on Figure 2. Assuming one additional nonviolent failure (shaded area) is likely to render the results of Model 7 statistically insignificant (majority of $p > 0.1$).

The above simulation holds all other aspects of the original analysis constant, allowing for a straightforward interpretation of the results. While there may be other possible adjustments to the underlying data, the variable selection, and modeling choices, they would lead this inquiry farther away from the original study and are, ultimately, beyond the scope of this research note. That being said, one issue related to variable selection that is of particular theoretical relevance is the role of campaign size. As indicated above, the size of a campaign is a key mechanism through which civil resistance affects campaign outcomes. This would render the covariate that conditions on the peak number of campaign participants (*# participants, log*) a posttreatment variable. Therefore, the coefficient estimate of *nonviolent campaign*, which is at the center of the stress tests presented above, may not represent the total effect of nonviolent resistance. Conditioning on the number of participants likely partials out part of a theoretical mechanism and result in various biases of

12. The focus on statistical significance follows WCRW's use of *p*-values for null-hypothesis significance testing (NHST). Appendix section A.4 shifts attention to the distributions of the coefficient estimates.

unknown direction (Dworschak 2023). In the appendix section A.5, I discuss the merits and pitfalls of including *number of participants* as a covariate in the analysis. Deviating from the original analyses in WCRW, I exclude *number of participants* and repeat the above simulations. While this increases the coefficient estimate of *nonviolent campaign* and makes it more robust to the inclusion of additional failed nonviolent campaigns, I also find that it may not be robust to the presence of omitted variable bias.

4 Discussion: Implications for the study of civil resistance effectiveness

The seminal work "Why Civil Resistance Works" (WCRW; Chenoweth and Stephan 2011) serves as the main reference for the notion of civil resistance success. I reproduce the core findings of WCRW and find that the effect of nonviolent resistance on campaign success is insignificant in three of the eight main models, suggesting a relevant degree of model dependence. This indicates uncertainty regarding WCRW's evidence on civil resistance effectiveness, even before subjecting the findings to additional stress tests. I then discuss the potential for campaigns having been overlooked, how this may affect the findings, and subject them to stress tests that quantify their sensitivity to sample selection. Simulating unobserved campaigns in line with the hypothesized data generating process of the missingness, I find that the remaining results turn insignificant when introducing one additional campaign to the data. When using a bootstrapping approach, the effect of civil resistance on campaign success is not statistically distinguishable from zero already before accounting for unobserved failed nonviolent campaigns. When accounting for the ambiguous role of campaign size in the model, the results are more robust to the introduction of missing campaigns, but also vulnerable to omitted variable bias. In conclusion, WCRW does not provide robust statistical evidence for the finding that civil resistance was more effective than violent resistance.

This does not mean that civil resistance does "not work" or that violent strategies are more effective. At most, these results may be seen as indicative of there being no difference in effectiveness between violent and nonviolent approaches, making civil resistance a viable alternative to violence (Helvey 2004; Sharp 2003, 1973). Moreover, effectiveness of civil resistance in achieving immediate policy goals is different from effectiveness in inducing long-term change, like facilitating transition processes to sustained democracy (Chenoweth 2021; Pinckney 2020; Lambach et al. 2020; Bayer, Bethke and Lambach 2016; Celestino and Gleditsch 2013; Chenoweth and Stephan 2011). In addition, and as I will discuss below, the aggregate nature of the replicated analyses, both conceptually and in their unit of observation, may impede detection of relevant patterns (cf., Chenoweth, Pinckney and Lewis 2018). Last but not least, there remains a rich body of qualitative work on civil resistance success (Nepstad 2015; Chenoweth and Stephan 2011; Nepstad 2011;

Schock 2005; Sharp 1973). Therefore, this research note casts doubt on the general claim that civil resistance is more effective than violent resistance only as far as the role of WCRW's quantitative findings goes in establishing said claim.

However, this role is profound. At the time of this writing, there is no study on the general outcome of civil resistance success that would be of comparable cross-national breadth and analytical depth, and WCRW's influence on scholarly debate and public discourse can hardly be overstated. This impact amplifies the relevance of continuously reviewing and discussing its evidence. Providing an ultimate answer to the question of civil resistance effectiveness is beyond the scope of this research note. Instead, it serves to revive the question by demonstrating that the chief reference used to back up civil resistance success does not yield robust statistical evidence. WCRW signifies a groundbreaking effort published at the analytical forefront of peace studies in 2011. Since then, important advances in data availability and a growing awareness of observational causal inference enable to address some of the concerns that give rise to the apparent uncertainty surrounding WCRW's core findings. I suggest three implications for the study of civil resistance effectiveness.

First, alternative data sources may be less prone to systematic missingness and conceptual ambiguity. While there is never a "complete" source that yields information on the census of all resistance activity, it is important to mitigate a systematic undercount of nonviolent campaigns. An alternative to NAVCO's campaign-level data is presented by Hellmeier and Bernhard (2022), drawing on the recent v11 release of V-Dem that incorporates information on mass mobilization going back to 1900 (Coppedge et al. 2021). While these data currently lack information on the degree of violence involved in mass mobilization, they constitute an important source for levels of mass mobilization pre 1990s by setting new standards for top-down data collection and processing, thus boosting expert data reliability (cf. Type C coding; Pemstein et al. 2020).

Second, choices regarding the research design and statistical modeling matter for observational causal inference. To this end, research needs to pay attention to the underlying causal processes and map the treatment assignment mechanism of nonviolent resistance. As the field of peace studies is moving beyond regression models replete with avoidable biases (Dworschak 2023; Westreich and Greenland 2013; Achen 2005), future research on civil resistance success can better conceptualize and address endogenous dynamics. Another promising trend is the disaggregation of key concepts, like campaign outcomes and methods of resistance, allowing for a more nuanced theoretical framework and empirical assessment.

Third, public discourse needs to adapt to the wealth of new findings that render the popular image of civil resistance success more nuanced (Hellmeier and Bernhard 2022; Manekin and Mitts 2022; Gleditsch et al. 2021). Instead of painting civil resistance effectiveness in broad strokes, academic references and policy debates need to adapt to the increasing level of uncertainty cast by a better understanding of multifaceted causal mechanisms, data limitations, and deeply endogenous processes. While perceptions of general success thus find more limited empirical support, civil resistance' beneficial effect on more specific outcomes, like its ability to facilitate democratization and foster more durable democratic institutions, is better documented (Pinckney 2020; Lambach et al. 2020; Orazani and Leidner 2019; Kim and Kroeger 2019). Taking a step back and reviewing the most-cited comparative evidence on civil resistance success, this research note adds to this important discussion, corroborating that an unconditional narrative of civil resistance effectiveness is unlikely to be tenable.

References

Achen, Christopher H. 2005. ‘Let’s Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong’. *Conflict Management and Peace Science* 22 (4): 327–339.

Anisin, Alexei. 2020. ‘Debunking the Myths Behind Nonviolent Civil Resistance’. *Critical Sociology* 46 (7-8): 1121–1139.

———. 2021. ‘Reinforcing Criticisms of Civil Resistance: A Response to Onken, Shemia-Goeke, and Martin’. *Critical Sociology* 47 (7-8): 1205–1218.

Bayer, Markus, Felix S. Bethke and Daniel Lambach. 2016. ‘The democratic dividend of nonviolent resistance’. *Journal of Peace Research* 53 (6): 758–771.

Celestino, Mauricio Rivera, and Kristian Skrede Gleditsch. 2013. ‘Fresh carnations or all thorn, no rose? Nonviolent campaigns and transitions in autocracies’. *Journal of Peace Research* 50 (3): 385–400.

Chenoweth, Erica. 2016. ‘Letters to the Editors’. *Comparative Politics* 48 (4): 580–582.

———. 2020. ‘The Future of Nonviolent Resistance’. *Journal of Democracy* 31 (3): 69–84.

———. 2021. *Civil Resistance: What Everyone Needs to Know.* New York: Oxford University Press.

Chenoweth, Erica, and Kathleen Gallagher Cunningham. 2013. ‘Understanding nonviolent resistance: An introduction’. *Journal of Peace Research* 50 (3): 271–276.

Chenoweth, Erica, Jonathan Pinckney and Orion Lewis. 2018. ‘Days of rage: Introducing the NAVCO 3.0 dataset’. *Journal of Peace Research* 55 (4): 524–534.

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

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, Nazifa Alizada, David Altman et al. 2021. ‘V-Dem [Country–Year/Country–Date] Dataset v11.1’. *Varieties of Democracy (V-Dem) Project.*

Day, Joel, Jonathan Pinckney and Erica Chenoweth. 2015. ‘Collecting data on nonviolent action: Lessons learned and ways forward’. *Journal of Peace Research* 52 (1): 129–133.

Dworschak, Christoph. 2023. ‘Bias mitigation in empirical peace and conflict studies: A short primer on posttreatment variables’. Online pre-print. *Journal of Peace Research.*

Earl, Jennifer, Andrew Martin, John D. McCarthy and Sarah A. Soule. 2004. ‘The Use of Newspaper Data in the Study of Collective Action’. *Annual Review of Sociology* 30 (1): 65–80.

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer. 2015. 'The Next Generation of the Penn World Table'. (Version 10.) *American Economic Review* 105 (10): 3150–3182.

Gleditsch, Kristian Skrede, Marianne Dahl, Scott Gates and Belen Gonzalez. 2021. 'Accounting for Numbers: Group Characteristics and the Choice of Violent and Nonviolent Tactics'. *Economics of Peace and Security Journal* 16 (1).

Hellmeier, Sebastian, and Michael Bernhard. 2022. 'Mass Mobilization and Regime Change: Evidence From a New Measure of Mobilization for Democracy and Autocracy From 1900 to 2020'. *V-Dem Institute Working Paper Series* 2022 (128): 1–46.

Helvey, Robert L. 2004. *On Strategic Nonviolent Conflict: Thinking About the Fundamentals.* Boston, MA: The Albert Einstein Institution.

Honaker, James, Gary King and Matthew Blackwell. 2011. 'Amelia II: A Program for Missing Data'. *Journal of Statistical Software* 45 (7): 1–47.

Kim, Nam Kyu, and Alex M. Kroeger. 2019. 'Conquering and coercing: Nonviolent anti-regime protests and the pathways to democracy'. *Journal of Peace Research* 56 (5): 650–666.

King, Gary. 1995. 'Replication, Replication'. *PS: Political Science & Politics* 28 (3): 444–452.

King, Gary, James Honaker, Anne Joseph and Kenneth Scheve. 2001. 'Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation'. *American Political Science Review* 95 (1): 49–69.

Lambach, Daniel, Markus Bayer, Felix S. Bethke, Matteo Dressler and Véronique Dudouet. 2020. *Nonviolent resistance and democratic consolidation.* London: Palgrave Macmillan.

Lehoucq, Fabrice. 2015. 'Does Nonviolence Work?' *Comparative Politics* 48 (2): 269–287.

Manekin, Devorah, and Tamar Mitts. 2022. 'Effective for Whom? Ethnic Identity and Nonviolent Resistance'. *American Political Science Review* 116 (1): 161–180.

Nepstad, Sharon Erickson. 2011. *Nonviolent Revolutions: Civil Resistance in the Late 20th Century.* Oxford: Oxford University Press.

———. 2015. *Nonviolent Struggle: Theories, Strategies, and Dynamics.* Oxford: Oxford University Press.

Onken, Monika, Dalilah Shemia-Goeke and Brian Martin. 2021. 'Learning from Criticisms of Civil Resistance'. *Critical Sociology,* 08969205211025819.

Orazani, S. Nima, and Bernhard Leidner. 2019. 'The power of nonviolence: Confirming and explaining the success of nonviolent (rather than violent) political movements'. *European Journal of Social Psychology* 49 (4): 688–704.

Pemstein, Daniel, Kyle L. Marquardt, Eitan Tzelgov, Yi-ting Wang, Juraj Medzihorsky, Joshua Krusell, Farhad Miri and Johannes von Römer. 2020. 'The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data'. *V-Dem Institute Working Paper Series* 2020 (21): 1–30.

Pinckney, Jonathan. 2020. *From Dissent to Democracy.* New York: Oxford University Press.

Pischedda, Costantino. 2020. 'Ethnic Conflict and the Limits of Nonviolent Resistance'. *Security Studies* 29 (2): 362–391.

Schock, Kurt. 2005. *Unarmed Insurrections: People Power Movements In Nondemocracies.* Minneapolis: University Of Minnesota Press.

Sharp, Gene. 1973. *The Politics of Nonviolent Action.* Boston: Porter Sargent.

———. 2003. *There Are Realistic Alternatives.* Boston, MA: The Albert Einstein Institution.

Singer, J. David, Stuart Bremer and John Stuckey. 1972. 'Capability Distribution, Uncertainty, and Major Power War, 1820-1965'. In *Peace, War, and Numbers,* edited by Bruce Russett, 19–48. (COW version 6.0.) Beverly Hills: Sage.

Stephan, Maria J., and Erica Chenoweth. 2008. 'Why civil resistance works: The strategic logic of nonviolent conflict'. *International Security* 33 (1): 7–42.

Westreich, Daniel, and Sander Greenland. 2013. 'The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients'. *American Journal of Epidemiology* 177 (4): 292–298.

A Online Appendix

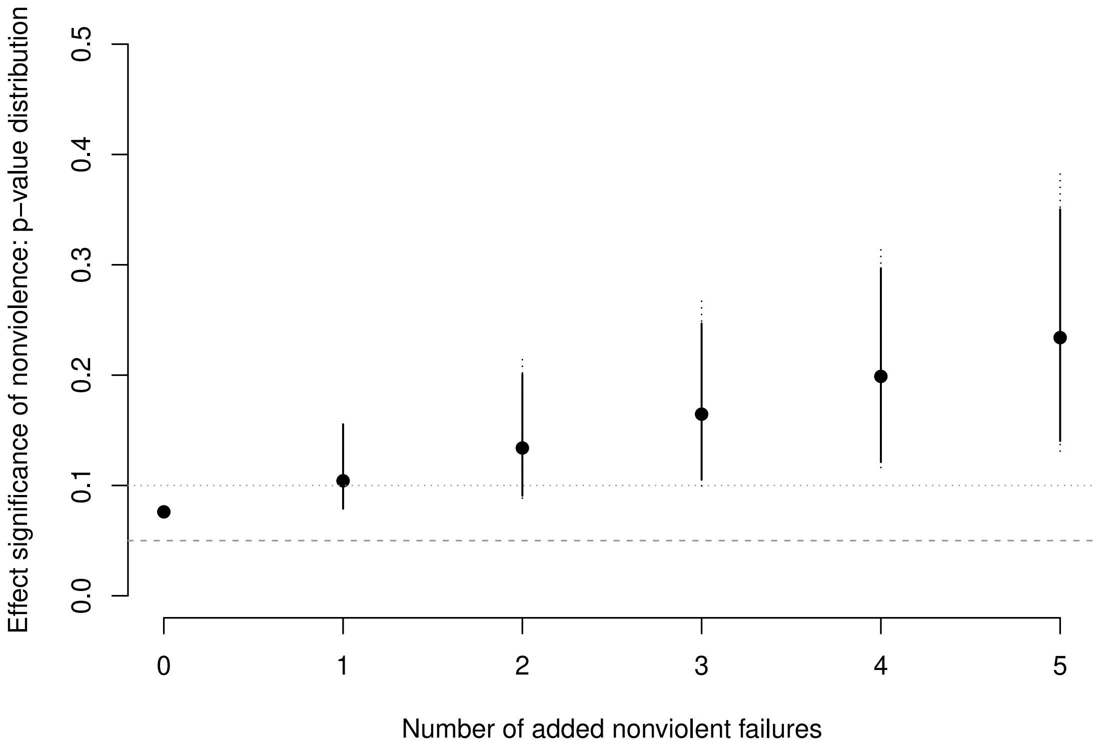
A.1 Supplementary tables and figures

Table A.1: R and STATA equivalence, Models 7 & 8 in WCRW's Table 3.1

	Dependent variable:					
	Campaign success					
	Model 7 (Original)	Model 7 (Repl.) (STATA)	Model 7 (Repl.) (R)	Model 8 (Original)	Model 8 (Repl.) (STATA)	Model 8 (Repl.) (R)
Nonviolent campaign	0.96*	0.963*	0.963*	0.43***	0.433	0.433
	(0.53)	(0.528)	(0.528)	(0.68)	(0.684)	(0.684)
Target polity score	0.00	0.000	0.0001	0.03	0.030	0.030
	(0.03)	(0.034)	(0.034)	(0.04)	(0.037)	(0.037)
Number of participants, log	0.39***	0.387***	0.387***	0.52***	0.520***	0.520***
	(0.12)	(0.123)	(0.123)	(0.17)	(0.166)	(0.166)
Population, log	–0.46***	–0.462***	–0.462**	–0.44**	–0.440**	–0.440**
	(0.18)	(0.178)	(0.178)	(0.21)	(0.209)	(0.209)
Target capabilities	1.63	1.628	1.628	3.88	3.879	3.879
	(5.64)	(5.638)	(5.638)	(7.52)	(7.523)	(7.523)
Violent regime repression	–1.78***	–1.777***	–1.777***	–2.77***	–2.766***	–2.766***
	(0.62)	(0.620)	(0.620)	(0.98)	(0.983)	(0.983)
Secessionist campaign	0.39	0.390	0.390	–0.34	–0.343	–0.343
	(1.34)	(1.336)	(1.336)	(1.35)	(1.346)	(1.346)
Anti-occupation campaign	2.69*	2.691*	2.691*	2.26*	2.257*	2.257*
	(1.41)	(1.411)	(1.411)	(1.23)	(1.230)	(1.230)
Regime change campaign	1.19	1.185	1.185	0.30	0.300	0.300
	(1.01)	(1.006)	(1.006)	(0.98)	(0.978)	(0.978)
Constant	0.003	0.027	0.027	0.75	0.749	0.749
	(1.87)	(1.875)	(1.875)	(2.20)	(2.202)	(2.201)
Observations	134	134	134	134	134	134
Continent fixed effects				✓	✓	✓
Decade fixed effects				✓	✓	✓
Cluster-robust SEs	✓	✓	✓	✓	✓	✓
Note:				*p<0.1; **p<0.05; ***p<0.01		

Full original models and full replication results of Model 7 and 8. Indicates equivalence of estimated standard errors between R and STATA, except a deviation due to rounding in the constant of Model 8.

Figure A.1: Results adjustment assuming unobserved nonviolent failures, updated data



Simulated *p*-value distributions using list-wise deleted observations padded with data from updated versions of WCRW's original sources. All other specifications correspond to Figure 2 on page 9 in the main text.

A.2 WCRW's quantitative evidence and the replication target

I focus my replication and subsequent assessment on WCRW's multivariate regression analyses (Table 3.1). The authors also present multiple auxiliary tests related to specific mechanisms. While these are insightful and suggestive, they are also more limited in their model specifications (e.g., simple bivariate cross-tabulation) and thereby less able to rule out alternative explanations. An instrumental variable regression (Chenoweth and Stephan 2011, 81, Table 3.3), which is used as a robustness check in support of the main findings in Table 3.1, is fully replicable. While the 2SLS model does not include the second-stage covariates in its first stage, correcting this oversight does not alter inference. However, among the variables that are employed as instruments are, e.g., the polity score and CINC score, which also serve as covariates in the main analysis. This may be considered at odds with their role as instruments and impedes further interpretation of the instrumental variable regressions, as the instruments' relationship with the outcome likely violates the exclusion restriction.

Given the research note's focus on WCRW, the replication and subsequent assessment of the role of missingness assume NAVCO 1.1 as baseline. Repeating the replication and assessment for NAVCO's data updates (Chenoweth and Lewis 2013; Chenoweth, Pinckney and Lewis 2018) and other data projects is a logical next step, but goes beyond the scope of this research note. Other datasets, including the NAVCO updates, differ from NAVCO 1.1 in their temporal and spatial coverage, their levels of aggregation, and they lack variables necessary for reproducing the multivariate analyses in WCRW. While these obstacles are not insurmountable (scope conditions can be redefined, information can be aggregated, and missing variables can be merged and proxied), they do put a limit on what can be achieved within the context of a single manuscript. Finding ways of meaningfully reconstructing WCRW's multivariate analyses with other data, whilst exercising due diligence at every step, means embarking on a whole new research endeavor in itself – even before being able to repeat any assessment of the role of missingness. Therefore, an exclusive focus on WCRW is a necessary scope condition of this research note and an important first step for future work. To the extent that NAVCO 1.1 may be seen as an "outdated" dataset, then, it begs the question of the extent to which the quantitative results presented in WCRW may be considered "outdated" and can still be referenced as evidence. This manuscript investigates this question. In the appendix section A.3 below, I briefly reflect on the sourcing and data generating process underlying NAVCO 1.1 and the other NAVCO updates. In the conclusion of the main text, I provide additional discussions of the topics of variable selection, model specification, and the aggregation of information.

A.3 A closer look at the WCRW source material

To better understand the potential for campaigns having been overlooked in WCRW's sample, it is useful to review the sources used to generate the data (cf. Chenoweth and Stephan 2011, 13, and the accompanying web appendix, 5-6). Information on nonviolent campaigns stem from multiple sources, which differ from the sources used to generate the sample of violent campaigns. Nonviolent campaigns were primarily drawn from the qualitative works of Karatnycky and Ackerman (2005), Schock (2005), and Carter, Clark and Randle (2006), which were then cross-checked with encyclopedias, case studies, and other sources found in Carter, Clark and Randle (2006), as well as "a dozen experts in nonviolent conflict" (Chenoweth and Stephan 2011 web appendix, 5). Importantly, none of the primary sources claim, or aim, to offer an exhaustive list of nonviolent campaigns: the study by Karatnycky and Ackerman (2005) is on political transitions from autocracy to democracy, thereby focusing on campaigns that would commonly be classified as "successful" (i.e., selecting on WCRW's dependent variable). Schock (2005) draws on a few selected successful and unsuccessful cases without a clearly defined sampling frame. Carter, Clark and Randle (2006, 5-6) define their sample as "transnational coverage of major examples of civil resistance and other significant nonviolent protest." While it is unclear what counts as "major examples" or "significant", they also explain that their study has "a particular emphasis on British examples" (6), and that sources were selected based on their availability in British libraries and exclude any non-English language material. Meanwhile, the source material for violent campaigns contains more systematic data collections, with better defined sampling frames: while the criteria for case selection in Sepp (2005) are as vague as those of the nonviolent source material, other sources on violent campaigns are the 2004 updated Correlates of War data project (Gleditsch 2004), the Clodfelter (2002) encyclopedia, and the insurgency data by Lyall and Wilson (2009).

In sum, there remains a higher ambiguity in the cases included, or left out, for nonviolent campaigns than for violent campaigns. This is not a shortcoming of WCRW, but a feature of data availability. As discussed in the main text, the risk of resistance campaigns not getting recorded is likely to be higher among those that remain unsuccessful and do not engage in violence – especially before the advent of social media.

This is not only the case for WCRW's NAVCO 1.1, but also for most of the NAVCO v1 and v2 series. While I cannot offer systematic replications of more NAVCO versions within the scope of this study,[1A] the others follow the same top-down procedure that generates a sampling frame of resistance campaigns based on academic expertise (case studies and expert consensus). As pointed out by Chenoweth, Pinckney and Lewis (2018) and Day, Pinckney and Chenoweth (2015),

1A. See also appendix section A.2 above for a discussion of these scope conditions.

these procedures are likely to yield a less complete sampling frame of maximalist campaigns than bottom-up data projects that systematically collect news reports on low-scale activities, like NAVCO 3.0. For more detail, see helpful discussions in the NAVCO data projects' supplementary materials. Supplementary material for NAVCO 1.3 is the first of the v1 series to mention the use of news articles for the creation of the candidate dataset, among other sources. While the details of this use of newswires are not clear (e.g., which search strings and what engine(s) were used, and for which time periods), this is a first and important step towards a more systematic and replicable sampling frame of campaigns. Event-level data also facilitate automated cross-referencing across multiple event data sources to further mitigate selection concerns (cf., Donnay et al. 2018). For generating inference on campaigns' successes or failures, and to apply relevant exclusion criteria like non-maximalist goals or campaign size, these event-level data can subsequently be aggregated to the campaign level (see, e.g., Pinckney 2016). That being said, event-level data also suffer form reporting bias (Weidmann 2016), and all data on dissent can only record realized resistance activity subject to strategic selection (Pierskalla 2010; Moore 1998; Lichbach 1987). Therefore, independent of which data is used, estimation needs to take censoring into account (cf. Ritter and Conrad 2016).

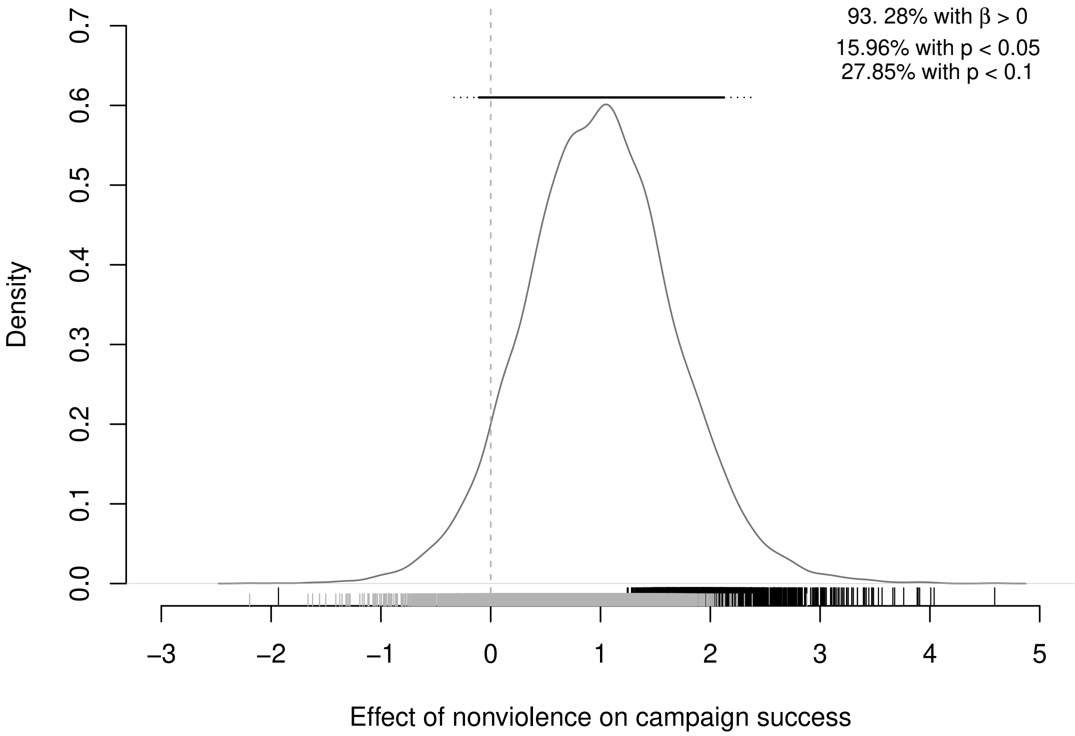
A.4 Adjusting sampling weights through bootstrapping

The sensitivity analysis in the main text adds hypothetical failed nonviolent campaigns to the original analysis sample, keeping the latter constant. This closely simulates the suggested data limitation of nonviolent failures having been overlooked, and thus are being added to the existing data. However, this approach also has important limitations. First, it impedes comparison to the original Model 7, because not only the base rates of success and failure change, but also the overall sample size changes. Second, it requires the introduction of previously unobserved covariate values. Third, the simulation keeps all 134 original observations of Model 7 constant, which limits insight into the role of data idiosyncrasies within the original sample and requires continued reliance on parametric assumptions for estimator uncertainty (standard errors and *p*-values). In other words, by analyzing the effect of nonviolence on campaign success based on one possible realization of the original dataset, the above simulation limits uncertainty to the sample variance and the few added hypothetical campaigns. Extreme values or sample idiosyncrasies in the original analysis sample may still skew the results in either direction.

Therefore, as a robustness check to the previous approach, I bootstrap the original analysis data while incrementally adjusting the sampling weights to over-sample failed nonviolent campaigns.[2A] It keeps the total number of observations constant (n = 134) and does not require multiple

2A. This is a version of a structured permutation test to correct for under-representation as suggested by Neumayer and Plümper (2017).

Figure A.2: Bootstrapped coefficients, original sample

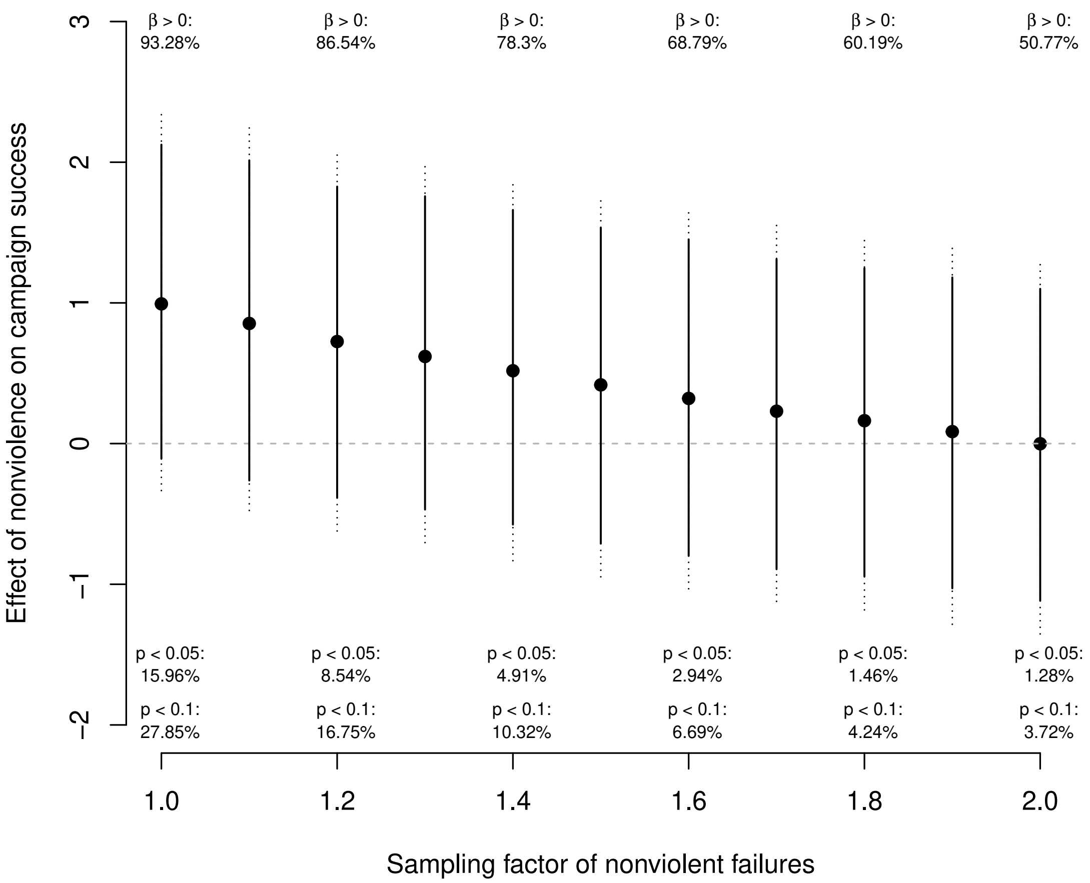


Simulated coefficient (β) distribution of *nonviolent campaign* based on 10,000 bootstrap samples. The solid and dotted line at the top represent 95% and 90% confidence intervals, respectively. The share of coefficients that pass conventional significance thresholds (based on parametric cluster-robust standard errors) are indicated on the right. The black rug represents the coefficients that pass the 0.05 threshold, the gray rug those that do not.

imputation of missing information for hypothetical nonviolent failures. Due to all observations being sampled with replacement, it also allows for a meaningful display of the full coefficient distribution as non-parametric representation of uncertainty. Figure A.2 shows the distribution of coefficient estimates for *nonviolent campaign* based on 10,000 bootstrap samples, assuming equal sampling weights. 93% of sampled coefficient estimates indicate a positive effect of nonviolent resistance methods on campaign success, while 7% indicate a negative effect. In this context, the 95% and 90% quantiles of the empirical distribution of coefficients can be interpreted like frequentist confidence intervals (CI),[3A] which are displayed at the top of the graph. Both the 95% CI (dotted line) and 90% CI (solid line) overlap with zero. In frequentist terms this suggests that, based on the original analysis sample and specifications of Model 7, the effect of *nonviolent campaign* on *campaign success* is not statistically distinguishable from zero. This is based on a standard bootstrap without having adjusted any sampling weights yet.

3A. Given an original sample size of 134 and a symmetric bootstrap distribution, I employ a simple percentile bootstrap (Hesterberg 2015; Davison and Hinkley 1997)

Figure A.3: Bootstrapped coefficients, sampling weight adjustment



Simulated coefficient (β) distributions of *nonviolent campaign* based on 10,000 bootstrap samples with varying sampling weights. The points represent the distribution means. The solid and dotted lines represent 95% and 90% confidence intervals, respectively. The share of coefficients that pass conventional significance thresholds (based on parametric cluster-robust standard errors) are indicated at the bottom.

Table A.2: Median numbers of sampled observations

	Violent campaigns		Nonviolent campaigns	
Sampling factor	Failure	Success	Failure	Success
(Original sample)	61	23	21	29
1.0	61	23	21	29
1.2	59	22	24	28
1.4	57	21	28	27
1.6	56	21	31	27
1.8	54	20	34	26
2.0	53	20	36	25

With these initial results not yielding a statistically significant effect of *nonviolent campaign*, it is not surprising that the proposed stress test in which the sampling weights are adjusted does not change this conclusion. As expected, accommodating the suggested unobserved nonviolent failures only attenuates the effect further. Figure A.3 shows the respective coefficient distributions. From left to right, they start with a distribution without weight adjustment (corresponds to Figure A.2), and proceed to step-wise increase the probability of drawing nonviolent campaigns up to a factor of 2. While this last step sees the distribution almost centered on zero, it also constitutes a fairly extreme scenario that assumes it to be twice as likely to overlook nonviolent failures than it is to overlook nonviolent success or violent campaigns. The median sample base rates for each step are listed in Table A.2 on page A-7. Just as the first simulation approach, these results assume that the true unobserved failed nonviolent campaigns are not systematically different from the observed nonviolent failures.

A.5 Campaign size as mediator and confounder

The replication and sensitivity analysis in the main text are based on the original model specifications as they appear in WCRW. While there may be many reasons for the inconsistent findings across models in the original replication, and for the sensitivity of Model 7 to unobserved nonviolent failures and nonparametric uncertainty estimates, one particular aspect warrants special attention: the role of the peak number of campaign participants as covariate in some of the models, including Model 7 and Model 8 (denoted as: *# participants, log*). As summarized above, campaign size is the main theoretical mechanism linking nonviolence to campaign success (see also, e.g., Nepstad (2011), DeNardo (1985) and Sharp (1973)). This makes the number of campaign participants, at least in part, a posttreatment variable. Conditioning on the number of participants may partial out part of the main theoretical mechanism, or induce related biases of unknown direction (Dworschak 2023). This does not necessarily mean that the *number of participants* should be omitted as a covariate from the models: it is also a confounder, in that nonviolent campaigns are more reliant on large numbers for their activities than violent ones. Moreover, campaign size might proxy for relevant unobserved pretreatment confounders (Cinelli, Forney and Pearl 2020; Angrist and Pischke 2009). Importantly, excluding it from the models is unlikely to bound the true effect of *nonviolent campaign* (Dworschak 2023). This highlights the difficulty of isolating and interpreting relevant patterns based on observational data.

In summary, depending on the role of *number of participants* in the data generating process, it functions as either a confounder or a mediator, or, most likely, as both. Therefore, both including and omitting *number of participants* biases the total effect estimate of *nonviolent campaign*: its omission as a covariate from the model may induce omitted variable bias, while its inclusion risks posttreatment bias and collider-stratification bias. The role of *number of participants* as main theoretical mechanism relaying the positive effect of civil resistance on campaign success makes it likely that its inclusion in a model might attenuate the total effect estimate of *nonviolent campaign*, although this assumes there are no strong negative second-hand confounders. This assumption cannot be tested by dropping *number of participants* from the analyses, because any change in the effect estimate of *nonviolent campaign* may be distorted by omitted variable bias: for example, a larger *number of participants* may both increase a campaign's success rate, as well as enable it to adopt nonviolent means (Gleditsch et al. 2021).

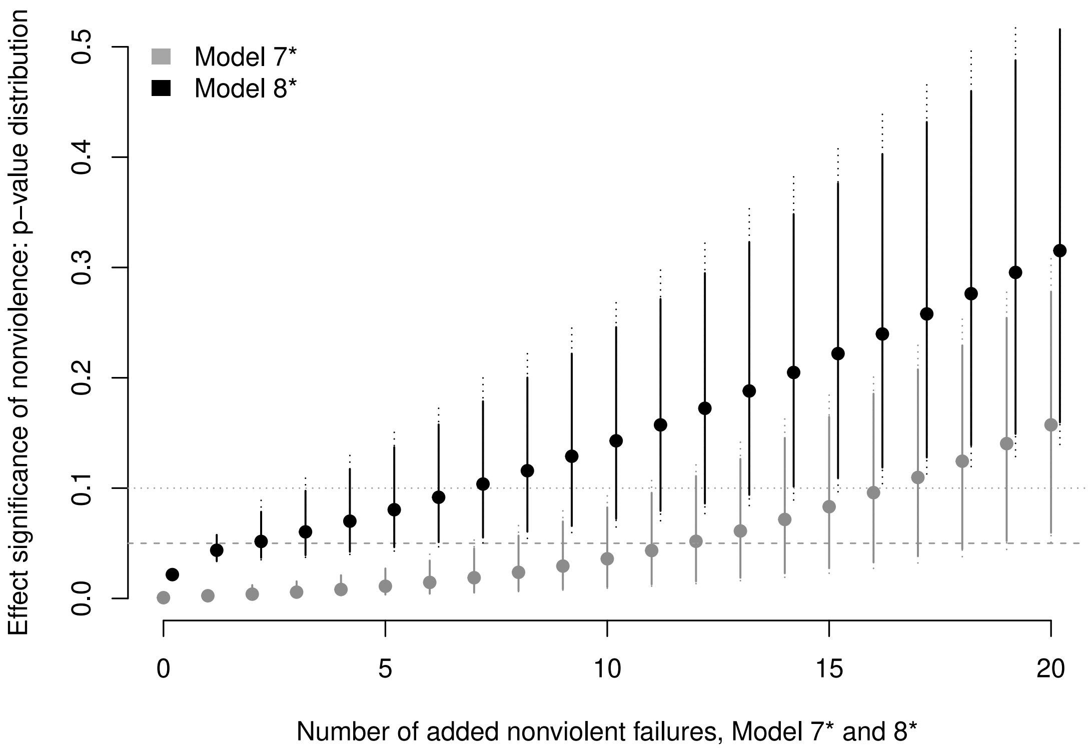
Table A.3: Main models without conditioning on campaign size

	Dependent variable:			
	Campaign success			
	(7)	(7*)	(8)	(8*)
Nonviolent camp.	0.963*	1.620***	0.433	1.339**
	(0.528)	(0.490)	(0.684)	(0.595)
Target polity score	✓	✓	✓	✓
# participants, log	✓		✓	
Population, log	✓	✓	✓	✓
Target capabilities	✓	✓	✓	✓
Violent reg. repress.	✓	✓	✓	✓
Secessionist camp.	✓	✓	✓	✓
Anti-occupat. camp.	✓	✓	✓	✓
Reg. change camp.	✓	✓	✓	✓
Continent FEs			✓	✓
Decade FEs			✓	✓
Observations	134	134	134	134
Cluster-rob. SEs	✓	✓	✓	✓
Note:		*p<0.1; **p<0.05; ***p<0.01		

The effect estimate of *nonviolent campaign* with and without (marked with *) campaign size (abbreviated as: *# participants, log*). Models 7 and 8 mirror the results in Table 1 and are included to facilitate comparison.

To explore this issue, I exclude *number of participants* from the analyses, thereby avoiding posttreatment and collider-stratification bias at the cost of accepting omitted variable bias. I re-run the previous simulations on these reduced models, and conclude by quantifying their sensitivity to omitted variable bias. As before, I focus on the most comprehensive model specifications to rule out more alternative explanations. Table A.3 shows Model 7 and Model 8 with and without (marked with *) *number of participants*. There is a general increase in effect size and significance level of *nonviolent campaign* when excluding *number of participants*, which, as explained above, may be due to various reasons.

Figure A.4: Assuming unobserved nonviolent failures, omitting campaign size

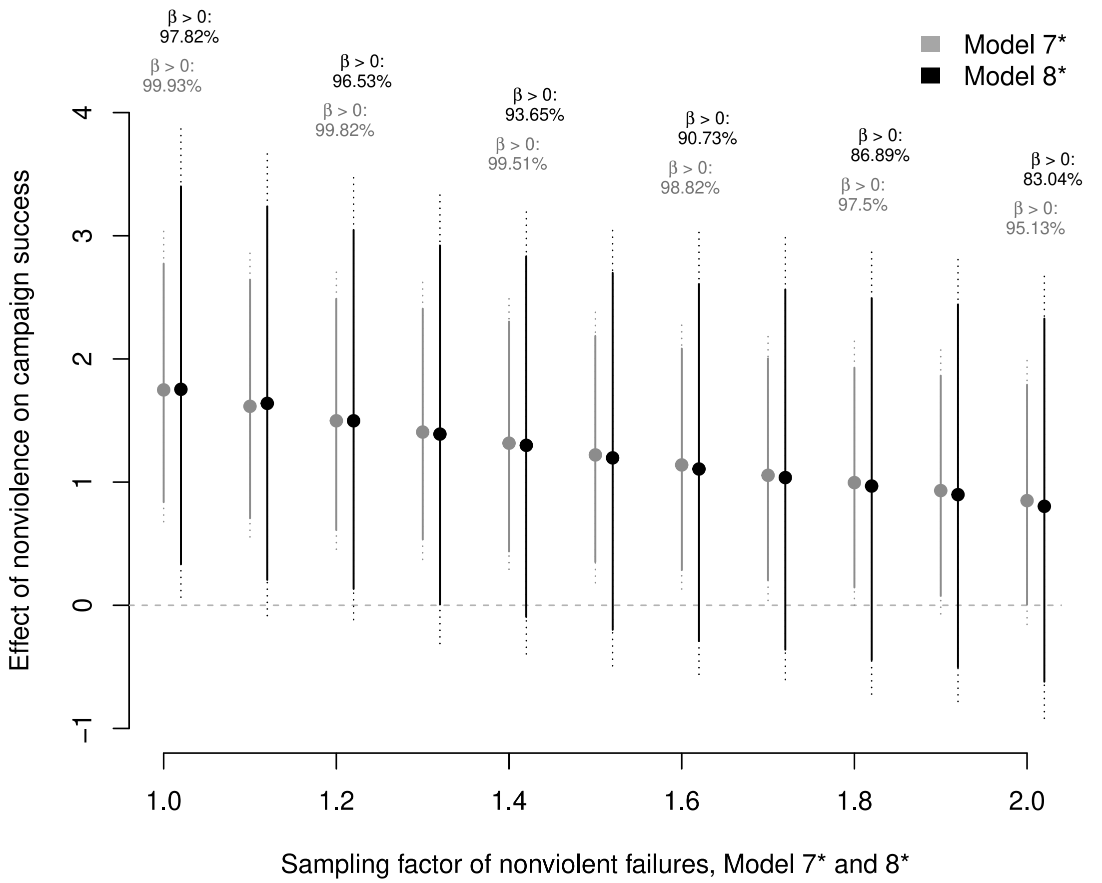


Simulated *p*-value distributions of the effect of *nonviolent campaign* while not conditioning on campaign size, based on Model 7* and Model 8* in Table A.3. Simulation specifications correspond to Figure 2 above.

Figures A.4 and A.5 visualize the simulation results based on Model 7* and Model 8*. In line with the increased magnitude of the effect estimate observed in Table A.3, not conditioning on *number of participants* also increases the effect estimates' robustness to unobserved failed nonviolent campaigns. The simulation approach in Figure A.4, adding hypothetical nonviolent failures to the original analysis sample, shows that a majority of possible scenarios yield a statistically significant effect estimate of *nonviolent campaign* when adding up to seven additional nonviolent failures based on Model 8*, and up to 17 based on Model 7*. Similarly, Figure A.5 indicates that failed nonviolent campaigns need to be oversampled by a factor of 1.4 to overturn the significant effect estimate of Model 8*, and a factor of over 2 in the case of Model 7*.

In sum, the effect estimate of *nonviolent campaign* is both substantially and statistically more significant, and is more robust to unobserved failed nonviolent campaigns, when *number of participants* is not included in the analysis. As discussed above, this increase is likely due to a combination of *number of participants* partialing out part of the effect of *nonviolent campaign* on *campaign success* when it is included in the models, as well as an inflation due to confounding when it is not included in the models. This confounding influence could mean that the observed positive

Figure A.5: Sampling weight adjustment, omitting campaign size



Simulated coefficient (β) distributions of *nonviolent campaign* while not conditioning on campaign size, based on Model 7* and Model 8* in Table A.3. Simulation specifications correspond to Figure A.3 above.

effect of *nonviolent campaign* on *campaign success* is an artifact of omitted variable bias rather than a genuine indication of civil resistance effectiveness. Therefore, to understand how sensitive the new effect estimate is to such omitted variable bias, I use a computational sensitivity analysis that simulates the influence of this bias based on observed exogenous covariates.

To determine whether accounting for a hypothetical ex ante measure of *number of participants* in Model 7* and Model 8* would likely overturn the effect of *nonviolent campaign*, I use *sensemakr* to examine the sensitivity of the effect of *nonviolent campaign* to omitted variable bias (Cinelli and Hazlett 2020; Cinelli, Ferwerda and Hazlett 2020). First, I re-estimate the models of Table A.3 as linear probability models (LPM). A linear probability model provides an unbiased treatment effect estimate for *nonviolent campaign* while facilitating the sensitivity analysis and enabling

Table A.4: Sensitivity analysis using linear probability models

	Dependent variable:			
	Campaign success			
	(7 LPM)	(7* LPM)	(8 LPM)	(8* LPM)
Nonviolent camp.	0.204**	0.336***	0.116	0.258**
	(0.101)	(0.090)	(0.111)	(0.103)
Target polity score	✓	✓	✓	✓
# participants, log	✓		✓	
Population, log	✓	✓	✓	✓
Target capabilities	✓	✓	✓	✓
Violent reg. repress.	✓	✓	✓	✓
Secessionist camp.	✓	✓	✓	✓
Anti-occupat. camp.	✓	✓	✓	✓
Reg. change camp.	✓	✓	✓	✓
Continent FEs			✓	✓
Decade FEs			✓	✓
Observations	134	134	134	134

Note: *p<0.1; **p<0.05; ***p<0.01

Estimates based on linear probability models (LPM), mirroring Table A.3.

Table A.5: Analysis of sensitivity to omitted variable bias

Outcome: *Campaign success*							
Model	Treatment	df	Est.	S.E.	t-value	$R^2_{Y\sim D\|X}$	$RV_{q=1,\alpha=0.05}$
Model 7*	*Nonviolent camp.*	125	0.336	0.090	3.743	10.1%	14.5%
		Bound (1x campaigns' goals): $R^2_{Y\sim Z\|X,D}$ = 12%, $R^2_{D\sim Z\|X}$ = 17%					
Model 8*	*Nonviolent camp.*	115	0.258	0.103	2.499	5.2%	4.6%
		Bound (1x campaigns' goals): $R^2_{Y\sim Z\|X,D}$ = 10%, $R^2_{D\sim Z\|X}$ = 12%					

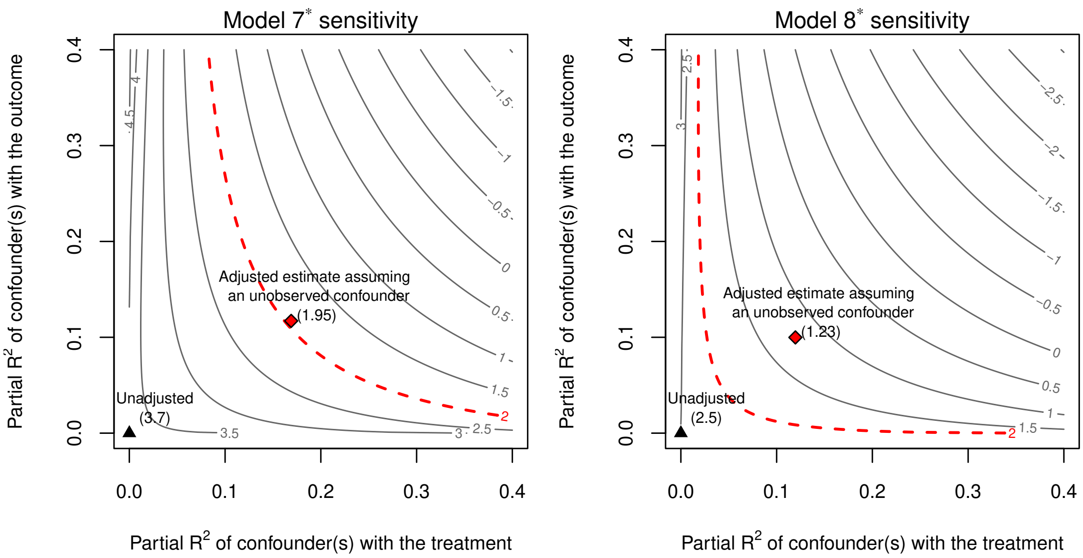
Analysis of sensitivity of Model 7* and Model 8* to omitted variable bias, as may be induced by en ex ante influence of *number of participants. Campaigns' goals* serves as a (more) exogenous benchmark covariate.

the use of *sensemakr*.[4A] Table A.4 shows the results of the LPMs. Second, I estimate the sensitivity of LPMs 7* and 8*. Table A.5 shows the omitted variable bias bounds for the two models, where the robustness value $RV_{q=1,\alpha=0.05}$ indicates the residual variance to be explained by a confounder in both the treatment and the outcome to render the treatment effect estimate statistically insignificant (p > 0.05). A confounder that explains less residual variance in the treatment must in turn explain

4A. There are no concerns over functional form considerations: even if the mapping of outcome values was of interest for the task at hand, the binary nature of the treatment reduces the test to a simple comparison of means.

more in the outcome, and vice versa, to lead to the same conclusion. Figure A.6 shows how this trade-off maps for both models, with the axes indicating the residual variances and the red dashed line representing the significance threshold for *nonviolent campaign*. In the bottom left corner is the original effect estimate for *nonviolent campaign* of Model 7* and Model 8*. Accounting for an hypothetical ex ante measure of *number of participants* would move the adjusted effect estimate further towards the top right, and thus further towards insignificance, depending on the residual variances explained by this confounder.

Figure A.6: Sensitivity analysis when omitting *number of participants*



Visualization based on Table A.5. The hypothetical unobserved confounder is simulated based on *campaigns' goals*. The "unadjusted" marker represents the original coefficient estimate as it appears in Table A.4, with the number in brackets showing the original *t*-value. The red "adjusted" marker represents the coefficient when accounting for the simulated omitted variable bias, with the adjusted *t*-value in brackets.

How likely is it that an exogenous confounder, like the ex ante influence of *number of participants*, would explain enough variance to turn the effect of *nonviolent campaign* insignificant? Due to the lack of such an ex ante measure of *number of participants*, I use the available information on *campaigns' goals* (secessionist, anti-occupation, or regime change) to quantify the robustness of the effect of *nonviolent campaign* to omitted variable bias. While *campaigns' goals* is similar to *number of participants* in that it constitutes an important determinant of both *campaign success* and the tactics a campaign adopts, it is more likely to be exogenous to the treatment assignment. Assuming an unobserved confounder of similar relevance as *campaigns' goals* that is orthogonal to the covariates moves the adjusted effect estimate of *nonviolent campaign* just above its significance threshold in Model 7*, and well above its significance threshold in Model 8*. This is visualized in

Figure A.6. In other words, if there was an unobserved confounder as strong as the benchmark covariate *campaign's goals*, it would fully account for the significant effect of *nonviolent campaign* in both models, even before introducing additional unobserved nonviolent failures.[5A]

In conclusion, when not conditioning on the *number of participants* in the analysis of civil resistance success, the findings on the effectiveness of nonviolence are mixed: the main effect estimates of *nonviolent campaign* increase and become more robust to unobserved nonviolent failures, but they are not robust to likely omitted variable bias. In other words, even when combining the effects of nonviolence and campaign size, their joint total effect exerts no statistically significant effect when taking confounding into account. However, there are important caveats to these results: first, *number of participants* is unlikely to be orthogonal to the other covariates. Second, the true magnitude of its confounding influence is unknown, as its partial R^2 with the treatment is a mixture of confounding, mediating, and collider effects. Nevertheless, these results were not yet combined with additional unobserved nonviolent failures, which would further contribute to their insignificance. Taken together, while the results of the sensitivity analyses are a useful approximation to better understand the role of campaign size in relation to the findings' overall robustness, they are also subject to uncertainty. They do highlight, however, that changing WCRW's model specifications to combine the effects of *nonviolent campaign* and *number of participants* does not suffice to ameliorate the lack of robustness suggested by the replication.

5A. The robustness values $RV_{q=1,\alpha=0.05}$ of all other models of WCRW Table 3.1, omitting *number of participants* from models 1-3, are of comparable magnitude, ranging between 11% and 18%. Applying stress tests of similar strength to these models, the effect of *nonviolent campaign* turns insignificant in models 2*, 3*, and 6*. In these models, however, the role of other alternative explanations is not determined.

Appendix references

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton: Princeton University Press.

Carter, April, Howard Clark and Michael Randle. 2006. *People Power and Protest Since 1945: A Bibliography of Nonviolent Action.* London: Housmans Bookshop Ltd.

Chenoweth, Erica, and Orion A. Lewis. 2013. 'Unpacking nonviolent campaigns: Introducing the NAVCO 2.0 dataset'. *Journal of Peace Research* 50 (3): 415–423.

Chenoweth, Erica, Jonathan Pinckney and Orion Lewis. 2018. 'Days of rage: Introducing the NAVCO 3.0 dataset'. *Journal of Peace Research* 55 (4): 524–534.

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

Cinelli, Carlos, Jeremy Ferwerda and Chad Hazlett. 2020. 'sensemakr: Sensitivity Analysis Tools for OLS in R and Stata'. Working paper. Accessed July 7, 2021. *ResearchGate Online.*

Cinelli, Carlos, Andrew Forney and Judea Pearl. 2020. 'A Crash Course in Good and Bad Controls'. Working paper. Accessed March 3, 2021. SSRN doi 10.2139/ssrn.3689437.

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

Clodfelter, Micheal. 2002. *Warfare and Armed Conflicts: A Statistical Reference to Casualty and Other Figures, 1500-2000.* Jefferson, NC, USA: McFarland.

Davison, A. C., and D. V. Hinkley. 1997. *Bootstrap Methods and their Application.* Cambridge, UK: Cambridge University Press.

Day, Joel, Jonathan Pinckney and Erica Chenoweth. 2015. 'Collecting data on nonviolent action: Lessons learned and ways forward'. *Journal of Peace Research* 52 (1): 129–133.

DeNardo, James. 1985. *Power in Numbers.* (2016 reprint.) Princeton, NJ, USA: Princeton University Press.

Donnay, Karsten, Eric T. Dunford, Erin C. McGrath, David Backer and David E. Cunningham. 2018. 'Integrating Conflict Event Data'. *Journal of Conflict Resolution* 63 (5): 1337–1364.

Dworschak, Christoph. 2023. 'Bias mitigation in empirical peace and conflict studies: A short primer on posttreatment variables'. Online pre-print. *Journal of Peace Research.*

Gleditsch, Kristian. 2004. ‘A Revised List of Wars Between and Within Independent States, 1816-2002’. *International Interactions* 30 (3): 231–262.

Gleditsch, Kristian Skrede, Marianne Dahl, Scott Gates and Belen Gonzalez. 2021. ‘Accounting for Numbers: Group Characteristics and the Choice of Violent and Nonviolent Tactics’. *Economics of Peace and Security Journal* 16 (1).

Hesterberg, Tim C. 2015. ‘What Teachers Should Know About the Bootstrap: Resampling in the Undergraduate Statistics Curriculum’. *American Statistician* 69 (4): 371–386.

Karatnycky, Adrian, and Peter Ackerman. 2005. *How Freedom is Won: From Civic Resistance to Durable Democracy.* Freedom, CA, USA: Freedom House.

Lichbach, Mark Irving. 1987. ‘Deterrence or Escalation?: The Puzzle of Aggregate Studies of Repression and Dissent’. *Journal of Conflict Resolution* 31 (2): 266–297.

Lyall, Jason, and Isaiah Wilson. 2009. ‘Rage Against the Machines: Explaining Outcomes in Counterinsurgency Wars’. *International Organization* 63 (1): 67–106.

Moore, Will H. 1998. ‘Repression and Dissent: Substitution, Context, and Timing’. *American Journal of Political Science* 42 (3): 851–873.

Nepstad, Sharon Erickson. 2011. *Nonviolent Revolutions: Civil Resistance in the Late 20th Century.* Oxford: Oxford University Press.

Neumayer, Eric, and Thomas Plümper. 2017. *Robustness Tests for Quantitative Research.* Cambridge, England, UK: Cambridge University Press.

Pierskalla, Jan Henryk. 2010. ‘Protest, Deterrence, and Escalation: The Strategic Calculus of Government Repression’. *Journal of Conflict Resolution* 54 (1): 117–145.

Pinckney, Jonathan. 2016. *Making or Breaking Nonviolent Discipline in Civil Resistance Movements.* Washington, D.C.: ICNC Press.

Ritter, Emily Hencken, and Courtenay R. Conrad. 2016. ‘Preventing and Responding to Dissent: The Observational Challenges of Explaining Strategic Repression’. *American Political Science Review* 110 (1): 85–99.

Schock, Kurt. 2005. *Unarmed Insurrections: People Power Movements In Nondemocracies.* Minneapolis: University Of Minnesota Press.

Sepp, Kalev I. 2005. ‘Best Practices in Counterinsurgency’. *Military Review* 85 (3): 8–12.

Sharp, Gene. 1973. *The Politics of Nonviolent Action.* Boston: Porter Sargent.

Weidmann, Nils B. 2016. ‘A Closer Look at Reporting Bias in Conflict Event Data’. *American Journal of Political Science* 60 (1): 206–218.