Civil resistance in the streetlight:
Assessing comparative evidence on nonviolent effectiveness

Christoph Dworschak
Norwegian University of Science and Technology
christoph.dworschak@ntnu.no

10th February 2022

Work in progress; please do not share without the author's permission.

Does civil resistance work? Existing research emphasizes the effectiveness of nonviolent resistance movements over violent ones in achieving their 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 replicate WCRW's core results on nonviolent effectiveness. Second, I examine the data underlying this influential finding and discuss how cases may have been overlooked due to a streetlight effect. Third, I determine whether the conclusion of nonviolent effectiveness may be idiosyncratic to an incomplete sample or specific model specifications by quantifying the findings' sensitivity using simulations. I find that WCRW's main findings on nonviolent effectiveness lack statistical significance, and are not robust to bootstrapping, sample selection bias, and omitted variable bias. Reference to WCRW, and to its finding that nonviolent resistance is typically more effective than violent resistance, is prevalent across international relations, comparative politics, and mass media. Determining the robustness of this notion is relevant to practitioners and research spanning the past decade.

I thank Charles Butcher, Thea Johansen, and the participants of PRIO's Securing the Victory workshop for many helpful discussions and suggestions. I am indebted to Erica Chenoweth for their generous feedback and support. All mistakes are my own.

Civil resistance in the streetlight:
Assessing comparative evidence on nonviolent effectiveness

Does civil resistance work? Existing research emphasizes the effectiveness of nonviolent resistance movements over violent ones in achieving their 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 replicate WCRW's core results on nonviolent effectiveness. Second, I examine the data underlying this influential finding and discuss how cases may have been overlooked due to a streetlight effect. Third, I determine whether the conclusion of nonviolent effectiveness may be idiosyncratic to an incomplete sample or specific model specifications by quantifying the findings' sensitivity using simulations. I find that WCRW's main findings on nonviolent effectiveness lack statistical significance, and are not robust to bootstrapping, sample selection bias, and omitted variable bias. Reference to WCRW, and to its finding that nonviolent resistance is typically more effective than violent resistance, is prevalent across international relations, comparative politics, and mass media. Determining the robustness of this notion is relevant to practitioners and research spanning the past decade.

Word count: 8351

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; Ackerman and Kruegler 1994; Sharp 1973). 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; 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. Does civil resistance work?

First, I replicate the main 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, this discussion is followed by simulations in which I quantify how sensitive the results are to such missingness. I find that the book's main findings may be overturned, on average, by adding one nonviolent unsuccessful campaign to the original analysis sample. I also probe the replication results by allowing for the influence of a key mechanism, campaign size, to contribute to the total effect of nonviolent resistance. Across all analyses, I find that the existing evidence for nonviolent effectiveness provided in WCRW is not conclusive. Specifically, the findings lack statistical significance, and are not robust to bootstrapping, sample

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."

selection bias, and omitted variable bias. I conclude with a discussion of the implications for the study of civil resistance effectiveness.

Re-examining core evidence on civil resistance success by replicating and dissecting WCRW is an important contribution to the field of International Relations. WCRW is one of the most influential works in the whole of peace science. Winner of the American Political Science Association's prestigious Woodrow Wilson Prize, WCRW's release fundamentally shaped the civil resistance research agenda. 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.

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 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 encounter other sources of uncertainty, including variable selection, sample idiosyncrasies, and omitted variable bias.

2 The limelight: Replicating the success of civil resistance

WCRW argues that civil resistance tends to be more successful than violent resistance, and quantitatively tests this argument using cross-national campaign data in multivariate analyses. These analyses constitute the replication target. 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.

WCRW develops a comprehensive theoretical framework linking civil resistance to campaign success (Chenoweth and Stephan 2011, ch. 2). In brief, civil resistance is argued to attract higher levels of participation, because nonviolent action lowers physical, informational, moral, and commitment barriers compared to violent action. First, physical barriers are lower due to a diverse set of nonviolent tactics. There is a broad spectrum of nonviolent resistance activities, none of which necessitate weapons training and combat, making them physically accessible to a larger portion of the population. Second, nonviolent resistance is argued to have an informational advantage due to participation being easily observable. Violent campaigns, on the other hand, pursue more clandestine operations, making their levels of participation more difficult to observe. Finding safety in the crowd, people are more likely to participate in nonviolent movements where they can see large numbers of others participate. Third, people may be hesitant to use violence in pursuit of their goals. This moral quandary decreases the portion of the population willing to join a movement that employs violence. Finally, the level of commitment required for joining in a violent campaign is explained to be higher due to ex ante training costs (less ad hoc participation possible), recruit screening, a lower chance of staying anonymous, and a higher risk of physical harm.

The ability of civil resistance to attract a larger and more diverse support base increases its opportunities to drain the regime's capacity. Nonviolent action is argued to facilitate defections among the security forces, lead repression to be ineffective or even backfire, induce tactical innovation that enables the movement to outmaneuver the regime, and motivate international sanctions against the government and external support for the campaign. 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 regressions and finds support for civil resistance effectiveness. I focus my replication on these multivariate analyses, which constitute the primary quantitative evidence that WCRW brings to bear for establishing the link between civil resistance and campaign success. I do not replicate auxiliary tests related to specific mechanisms: while they are insightful and suggestive, they are also more limited in their model specifications (e.g., bivariate cross-tabulation), and thereby less able to rule out alternative explanations. Therefore, I concentrate on the main hypothesis tests based on multivariate regressions, as they appear printed in Table 3.1 (Chenoweth and Stephan 2011, 70-71).[3]

The quantitative analyses in WCRW are based on a dataset capturing the traits of 323 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). Some supplementary analyses, as well as Stephan and Chenoweth (2008), make a three-way distinction between campaigns being either successful, partially successful, or fail. The main analyses in Chenoweth and Stephan (2011) draws on a dichotomous success-failure indicator, pooling partial success and failure together. Given this binary nature of the outcome variable, WCRW uses logistic regressions.

WCRW tests whether campaign success can be explained via the campaign's means being either violent or nonviolent. Similar to the outcome variable, the explanatory variable of interest is a dichotomous violence-nonviolence indicator (*nonviolent campaign*). It captures whether a campaign used predominantly nonviolent methods in the context of civil disobedience, like sit-ins, protests, strikes, and boycotts, or whether it employed violence in the context of an insurgency or revolution, like shooting, bombing, kidnapping, and infrastructure damage. WCRW draws on a number of covariates to mitigate confounding and isolate the effect of civil resistance on campaign success. These are the target state's polity score, the peak number of campaign participants, the state's population size, the target regime's capabilities (CINC score), whether the regime tries to violently suppress the campaign, the campaign's goal, continent fixes effects, and decade fixed effects. WCRW employs heteroskedasticity-robust standard errors clustered at the country level.

3. 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. Among the variables that are employed as instruments are, e.g., polity score and CINC score, which also serve as covariates in the main analysis. While this may be considered at odds with their role as instruments, a discussion of the exclusion restriction goes beyond the scope of this manuscript.

Replication results

In the following I reproduce all eight models reported as WCRW's Table 3.1. In combination with several qualitative case studies, the results from these eight models 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.

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:	0.905^{*}	0.516	0.426	1.085^{***}	1.258^{***}	1.076^{***}	0.963^{*}	0.433
Nonviolent camp.	(0.485)	(0.428)	(0.426)	(0.247)	(0.256)	(0.280)	(0.528)	(0.684)
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.

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 (Chenoweth and Stephan 2011, 70-71). The second row shows the results generated based on the replication material accompanying Chenoweth and Stephan (2011).[4] The original results are fully reproducible, with the exception of the significance levels indicated for the main variable of interest, *nonviolent campaign*, in Model 2 (0.52***), Model 3 (0.43***), and Model 8 (0.43***).[5] In other words, the coefficients and standard errors of WCRW's main analyses are fully replicable, but not some of the indicated significance levels. This implies that the effect of nonviolence on campaign success is not robust to different model specifications.[6] This suggests that it is not clear whether choosing nonviolent means over violent ones improves the prospect for campaigns' success. This does not mean that civil resistance does *not* work, but merely that findings based on the analyses in WCRW are inconclusive. Importantly, the results obtained 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 civil resistance effectiveness to date. Does this evidence bear scrutiny? Chenoweth and Stephan (2011) and subsequent research identify important limitations to the Nonviolent and Violent Campaigns and Outcomes (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 (Lehoucq 2015; Day, Pinckney and Chenoweth 2015). Chenoweth and Stephan (2011) anticipate and discuss many of these limitations, and Chenoweth led an exemplary effort to

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 other 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. Footnote § on page 71 in WCRW suggests that a high significance level is printed next to the insignificant coefficients of *nonviolent campaign* due the significance of the linear combination of both *nonviolent campaign* and *number of participants*. I thank Erica Chenoweth for kindly confirming. 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 interpretation of *number of participants* as part of the effect of *nonviolent campaign*. I explore the role of *number of participants* in the appendix, the results of which are summarized further below.

6. 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.

continuously revise and update the data, leading up to NAVCO 2.0 (Chenoweth and Lewis 2013) and NAVCO 3.0 (Chenoweth, Pinckney and Lewis 2018). This makes NAVCO one of the most groundbreaking data projects in the field to date. Data limitations can have wide-ranging implications for the analysis results they generate, and despite the scholarly awareness of the drawbacks of NAVCO 1.1, 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, the aforementioned 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.

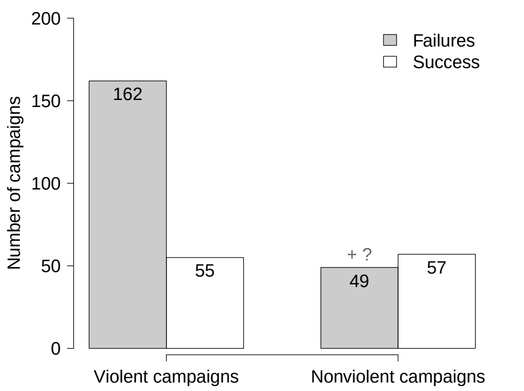
Violence in the streetlight

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 so-called streetlight effect: data availability determines which observations make it into the sample, which can lead to the exclusion of relevant cases. If this missingness is systematic, it can distort or nullify true population patterns, or produce spurious patterns that do not exist in the underlying population.

To understand the potential for campaigns having been overlooked in WCRW's sample, it is useful to first 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: violence is in the streetlight, and coverage of nonviolent resistance is generally lagging behind 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 the "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. Consequently, data on nonviol-

Figure 1: WCRW sample of campaigns



Empirical distribution of campaign success and resistance methods in WCRW. The gray questionmark indicates the concern for potential missingness of nonviolent failures.

ent campaigns is more likely skewed in favor of success: civil resistance activity with maximalist goals gets recorded when it reached 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). The empirical distribution of success and failure over violent and nonviolent resistance methods in WCRW's data is visualized in Figure 1, with the empirical implication of the theorized omission of nonviolent failures indicated by a question mark.

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. While the actual number of omitted cases cannot determined, this provides a qualitative baseline for the general concern that WCRW's sample may have omitted unsuccessful nonviolent campaigns. Whether an inclusion of these overlooked cases in WCRW's analyses may infringe upon its main findings, however, remains to be examined.

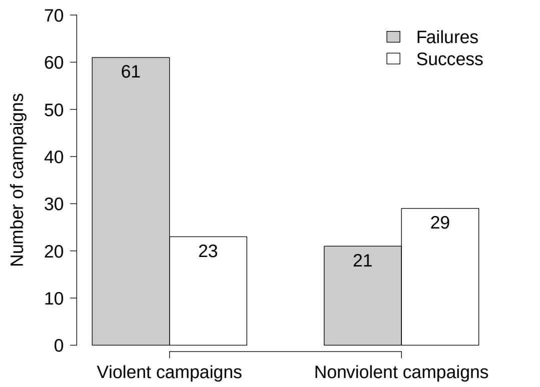
Chenoweth and Stephan (2011, 79, and web appendix, 8-9) offer a detailed discussion of the potential for nonviolent campaigns that did not mature enough to get recorded. They took several steps towards mitigating this risk, and conclude by simulating an extreme scenario in which half of the violent campaigns pose as failed nonviolent campaigns. Shifting the base rates shown in Figure 1 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 earlier. Therefore, below I present a systematic examination of the role of missingness for WCRW's main findings.

Simulating the effect of missingness

If there are 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 in two steps. First, I simulate failed nonviolent campaigns and add them to the original data, probing how many overlooked campaigns it would take to render the effect of nonviolence statistically indistinguishable from zero. Second, I turn to a more substantive inspection of how the results move when shifting baseline rates based on simulated coefficient distributions.

I use the original Model 7 (cf. Table 1 on page 7) as basis for these 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. As is the case for most other models, the analysis sample of Model 7 is a subsample of the full WCRW data due to the missingness in some of the covariates being addressed via list-wise deletion.[7] Figure 2 shows the relevant analysis sample.

Figure 2: WCRW analysis sample

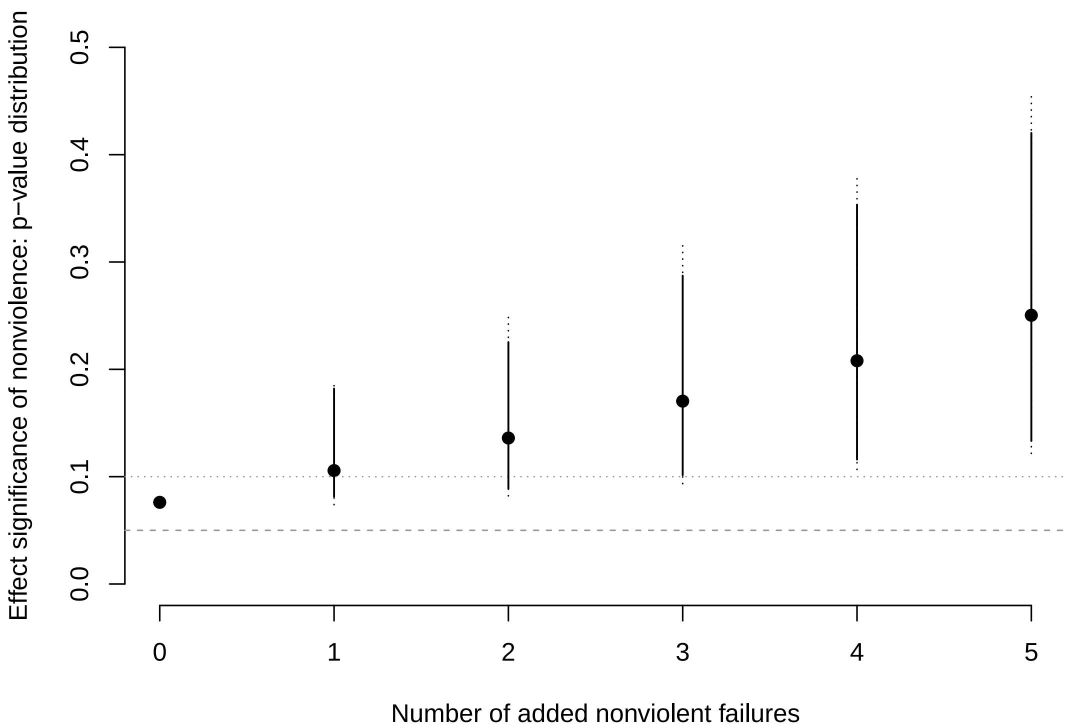


Empirical distribution of campaign success and resistance methods in WCRW's main analysis in Table 3.1, Model 7 and 8.

In a first step, I simulate additional failed nonviolent campaigns and add them to the analysis data. 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 (typically in *population size*). These omitted data include 28 failed nonviolent campaigns. I pad the missing cells using multiple imputation, simulating 50 different datasets of the 28 nonviolent failures. 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 its original sample, plus two randomly drawn campaigns of the 1400, giving as many possible analysis results.

7. While each of polity score, government capacity, and number of participants have a few missing observations, most missingness is introduced by an unknown country population size. As an additional check, I ran a replication of all models in which the sample is held constant (n = 134) across model specifications. This does not substantially change the results.

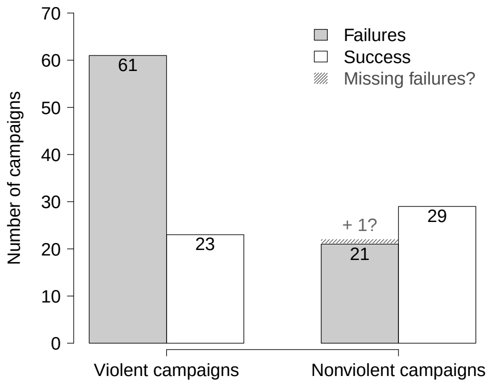
Figure 3: Effect adjustment assuming unobserved nonviolent failures



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.

The results of this sensitivity analysis are visualized in Figure 3. The left-most point shows the effect significance 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. This 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 would suffice for the effect of *nonviolent campaign* to not be statistically distinguishable from zero anymore.[8]

Figure 4: Sensitivity of existing evidence



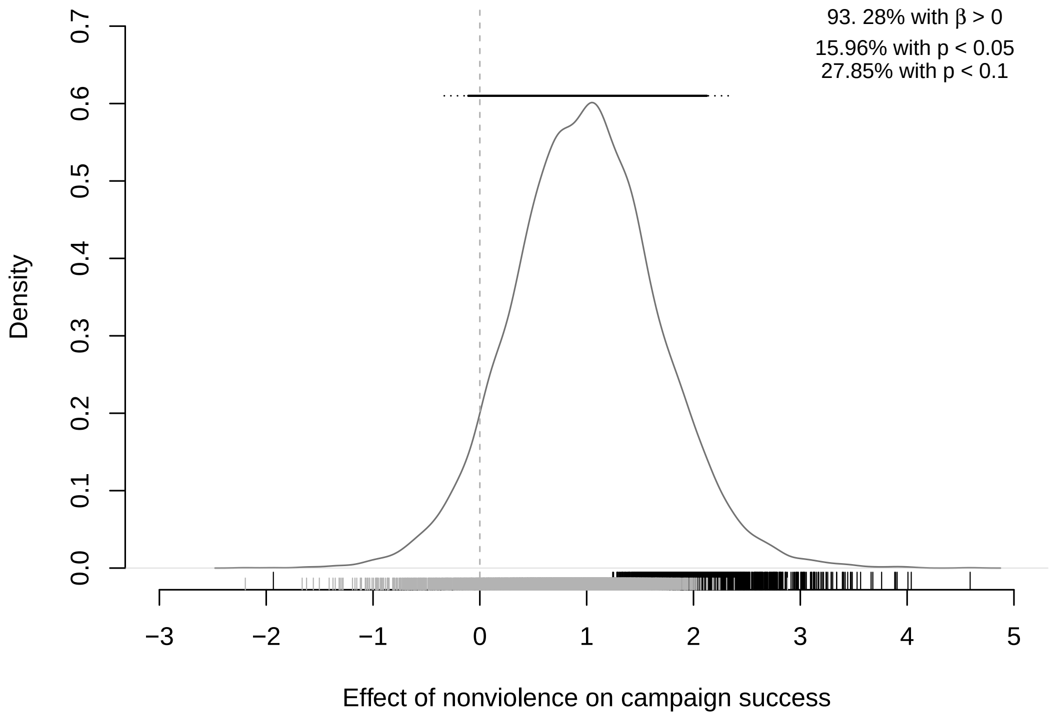
Based on Figure 3, assuming one additional nonviolent failure (shaded area) is likely to render the results of Model 7 statistically insignificant (majority of $p > 0.1$).

The above sensitivity analysis 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 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 favor failed nonviolent campaigns. This keeps the total number of observations constant (n = 134) and does not require multiple 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 instead of relying on the normality assumption

8. This finding assumes that the list-wise deleted nonviolent failures provide a reasonable baseline for assuming covariate values of unobserved nonviolent failures. 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 3 and are visualized in the Online Appendix, Figure A.1 on page A-2.

Figure 5: Bootstrapped coefficients, original sample

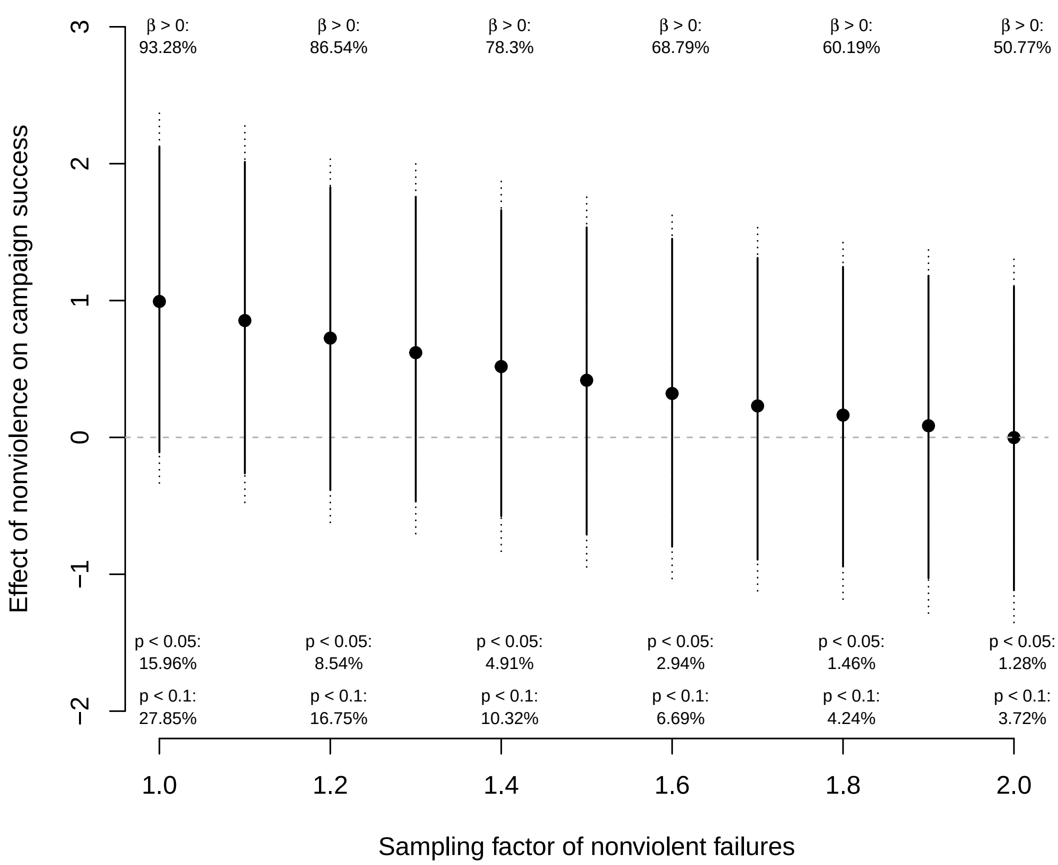


Simulated coefficient (β) distribution of *nonviolent campaign* based on 10,000 bootstrap samples. The solid and dotted line at the top represent 95% and 90% confidence intervals, respectively. The share of coefficients that pass conventional significance thresholds (based on parametric cluster-robust 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.

and *p*-values. Figure 5 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),[9] 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.

9. 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 6: 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.

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 6 shows the respective coefficient distributions. From left to right, they start with a distribution without weight adjustment (corresponds to Figure 5), 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 nonviolent success or violent campaigns. The median sample base rates for each step are listed in Table 2. 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.

Table 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

In conclusion, the finding that civil resistance is more effective than violent resistance, as it is presented in WCRW, is subject to uncertainty. The mere reproduction of results, based on WCRW replication material, suggests that the finding is heavily dependent on model specifications. Meanwhile, one of the two most comprehensive models, Model 7, continues to show a positive and statistically significant effect in this initial reproduction. Therefore, subsequent analyses concentrate on this model, assessing its sensitivity to unsuccessful nonviolent campaigns having been overlooked. A systematic underreporting of failed nonviolent campaigns is likely due to a streetlight effect (violence bias; Day, Pinckney and Chenoweth 2015) that leads to nonviolent campaigns being left-censored: those that do not garner a large-enough following or are successful before being repressed or before fading out do not get recorded. Based on the original WCRW data, I simulate these hypothetical unobserved nonviolent failures and examine how sensitive the effect of nonviolence on campaign outcomes is to their inclusion in the main model. I find that assuming one additional nonviolent failure is enough to turn a majority of potential results insignificant when using parametric uncertainty estimates. When relaxing this assumption, the effect of civil resistance on campaign success is not statistically distinguishable from zero already before accounting for unobserved failed nonviolent campaigns.

Finally, it may be argued that the coefficient estimate of *nonviolent campaign*, which is at the center of these stress tests, does not represent the total effect of nonviolent resistance. As indicated in the review of WCRW's theory and analysis, the size of a campaign is an important 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, conditioning on the number of participants may partial out part of a theoretical mechanism and result in various biases of unknown direction (Dworschak 2021). In the appendix, I discuss the merits and pitfalls of including *number of participants* as a covariate in the analysis. I also repeat the above simulations while excluding *number of participants*. While the coefficient estimate of *nonviolent campaign* gains in strength and is more robust to the inclusion of additional failed nonviolent campaigns, I also find that it is not robust to omitted variable bias, which is likely induced by the exclusion of *number of participants*. In sum, updating WCRW's main models to account for the theoretical link between *nonviolent campaign* and *number of participants*, and subjecting their combined effect to scrutiny, does not rectify the lack of robustness suggested by the replication and sensitivity analyses.

4 Discussion: Implications for the study of civil resistance effectiveness

Does civil resistance work? The seminal work "Why Civil Resistance Works" (WCRW; Chenoweth and Stephan 2011) serves as main reference for the notion of civil resistance effectiveness to date. I replicate the core findings of WCRW, discuss the potential for campaigns having been overlooked and how this may affect said findings, and subject them to stress tests that quantify their sensitivity to such sample selection. The original results of WCRW are fully replicable, with the exception of some of the significance levels of the main effect of interest. This indicates a relevant degree of uncertainty for the notion of civil resistance effectiveness, even before subjecting the findings to additional stress tests. Simulating unobserved campaigns in line with the hypothesized data generating process of the missingness, I find that the remaining findings turn insignificant when introducing one additional campaign to the data. Accounting for the ambiguous role of campaign size does not substantially change this inference.

This does not necessarily imply that civil resistance does not "work". Effectiveness of civil resistance in achieving their 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). 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). Moreover, there remains a rich body of qualitative work on civil resistance success (Nepstad 2015; Chenoweth and Stephan 2011; Nepstad 2011; Schock 2005; Sharp 1973). Thus, this research note casts doubt on the general claim that "civil resistance works" 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 thus WCRW's influence on scholarly debate and public discourse can hardly be overstated.[10] This impact amplifies the relevance of continuously reviewing and discussing its evidence. Past research has risen to the challenge, and among other important concerns it identifies systematic missingness as a likely issue (Anisin 2020; Chenoweth, Pinckney and Lewis 2018; Chenoweth 2016; Lehoucq 2015; Day, Pinckney and Chenoweth 2015). Connecting to this body of research, I show that such missingness would indeed infringe on some of WCRW's core findings, and thus on the main evidence for the notion that "civil resistance works" beyond individual case studies.

To bridge the gap between perceptions of nonviolent success and the limited comparative evidence, more research on civil resistance effectiveness is needed and perceptions need to be updated. WCRW signifies a groundbreaking effort published at the analytical forefront of peace studies. Meanwhile, important advances in data availability and a growing expertise in observational causal inference enable to address some of the concerns that give rise to the current 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 systematic missingness and conceptual ambiguity. Second, research designs and statistical modeling need to pay attention to underlying causal processes and map the treatment assignment mechanism of nonviolent resistance. Third, public discourse needs to adapt to the wealth of new findings that render the popular image of civil resistance success much more nuanced.

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 maximalist campaigns. Depending on the period of observation, a reliance on top-down academic and expert knowledge, rather than a systematic parsing of news reports for bottom-up information, is likely to exacerbate existing selection bias. As discussed above, the NAVCO v1 and v2 series follow such a top-down procedure, generating their sampling frame of nonviolent resistance based

10. Despite having looked long and hard, I'm still uncomfortable with this claim. Please let me know if you know quant cross-national studies with a "general" short-term outcome like "success".

on academic expertise (case studies and expert consensus).[11] As suggested by Chenoweth, Pinckney and Lewis (2018) and Day, Pinckney and Chenoweth (2015), this is 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. This also facilitates automated cross-referencing across multiple event data sources to further mitigate selection concerns (cf., Donnay et al. 2018). Finally, 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).

A major drawback of event data is its limited temporal scope. While the NAVCO data underlying WCRW goes back to 1900, coverage of major event datasets only starts in the 1990s, thus naturally limiting their inferential scope.[12] 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 make 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).

In addition to the quality of the underlying data, research design and modeling choices also matter for observational causal inference. As the field of peace studies is moving beyond regression models replete with avoidable biases (Dworschak 2021; Westreich and Greenland 2013; Achen 2005), and with the use of more fine-grained data, contemporary and future research on civil resistance success can better conceptualize and address endogenous dynamics, thus substantially improving identification. 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. A more differentiated approach to key concepts is especially important in the context of divergent trends that may otherwise conceal relevant pat-

11. For more detail, see helpful discussions in the NAVCO data projects' supplementary materials. Supplementary material for NAVCO 1.3 is the first of the series to mention the use of news articles for the creation of the candidate dataset, among other sources. While it is not clear how thorough the use of newswires is (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.

12. However, next to a revolution in source material, the advent of the internet also significantly altered patterns of contention (Jost et al. 2018; Christensen and Garfias 2018; Beissinger 2017; Zeitzoff 2017; Gohdes 2015; Youmans and York 2012) and repression (Lutscher et al. 2020; Roberts 2018; Asal et al. 2016). To the extent that this paradigm shift induces systematic change in key concepts and dynamics, it may render the usefulness of pre-1990s inference on resistance movements more limited.

terns, like pro-autocracy mobilization, fringe violence, and democratic backsliding (Hellmeier and Bernhard 2022; Chenoweth 2021, 2020).

This nuance, however, may also mean to relax assumptions and let go of prevalent notions of civil resistance "success". Instead of painting civil resistance effectiveness in broad strokes, discourse needs to adapt to the increasing level of uncertainty cast by multifaceted causal mechanisms, data limitations, and deeply endogenous dynamics. 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). It is more difficult to draw conclusions on civil resistance success in general Hellmeier and Bernhard (2022) and Manekin and Mitts (2022). 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.

Ackerman, Peter, and Christopher Kruegler. 1994. *Strategic Nonviolent Conflict: The Dynamics of People Power in the Twentieth Century.* Oxford, England, UK: Praeger.

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.

Asal, Victor, Jacob Mauslein, Amanda Murdie, Joseph Young, Ken Cousins and Chris Bronk. 2016. ‘Repression, Education, and Politically Motivated Cyberattacks’. *Journal of Global Security Studies* 1 (3): 235–247.

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

Beissinger, Mark R. 2017. ‘"Conventional" and "Virtual" Civil Societies in Autocratic Regimes on JSTOR’. *Comparative Politics* 49 (3): 351–371.

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

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, 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.

Christensen, Darin, and Francisco Garfias. 2018. 'Can You Hear Me Now? How Communication Technology Affects Protest and Repression'. *Quarterly journal of political science* 13 (1): 89.

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

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.*

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.

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. 2021. 'Bias mitigation in empirical peace and conflict studies: A short primer on posttreatment variables'. *Conflict research society conference paper.*

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

Gohdes, Anita R. 2015. ‘Pulling the plug: Network disruptions and violence in civil conflict’. *Journal of Peace Research* 52 (3): 352–367.

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.

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

Jost, John T., Pablo Barberá, Richard Bonneau, Melanie Langer, Megan Metzger, Jonathan Nagler, Joanna Sterling and Joshua A. Tucker. 2018. ‘How Social Media Facilitates Political Protest: Information, Motivation, and Social Networks’. *Political Psychology* 39 (S1): 85–118.

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

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.

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.

Lutscher, Philipp M., Nils B. Weidmann, Margaret E. Roberts, Mattijs Jonker, Alistair King and Alberto Dainotti. 2020. ‘At Home and Abroad: The Use of Denial-of-service Attacks during Elections in Nondemocratic Regimes’. *Journal of Conflict Resolution* 64 (2-3): 373–401.

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

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.

Nepstad, Sharon Erickson. 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. 2016. *Making or Breaking Nonviolent Discipline in Civil Resistance Movements.* Washington, D.C.: ICNC Press.

———. 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.

Roberts, Margaret E. 2018. *Censored: Distraction and Diversion Inside China's Great Firewall.* Princeton, NJ, USA: Princeton University Press.

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.

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.

Youmans, William Lafi, and Jillian C. York. 2012. 'Social Media and the Activist Toolkit: User Agreements, Corporate Interests, and the Information Infrastructure of Modern Social Movements'. *Journal of Communication* 62 (2): 315–329.

Zeitzoff, Thomas. 2017. 'How Social Media Is Changing Conflict on JSTOR'. *Journal of Conflict Resolution* 61 (9): 1970–1991.

A Online Appendix

1 Supplementary tables and figures

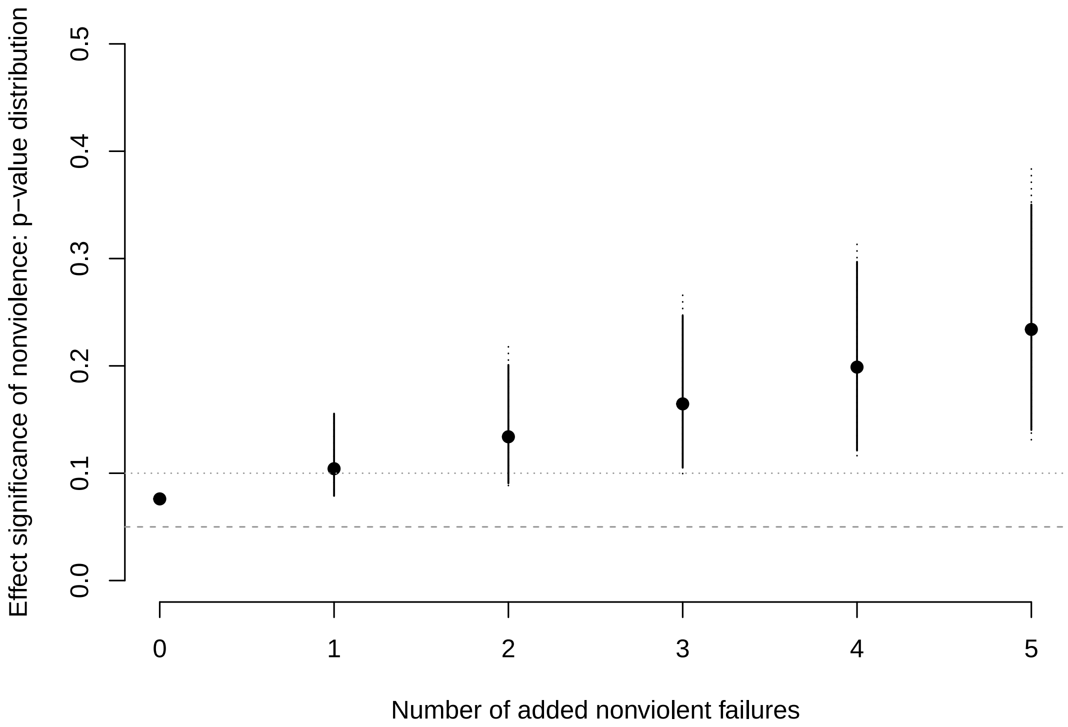
Table A.1: R and STATA equivalence, Models 7 & 8 in WCRW's Table 3.1

	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: Effect 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 3 on page 13 in the main text.

2 Campaign size as mediator and confounder

The replication and sensitivity analyses 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 2021). This does not necessarily mean that the *number of participants* should be omitted as a covariate from the models: it might proxy for important unobserved pretreatment confounders (Cinelli, Forney and Pearl 2020; Angrist and Pischke 2009), and excluding it from the models is unlikely to bound the true effect of *nonviolent campaign* (Dworschak 2021). However, it 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).

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

Table A.2: 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		

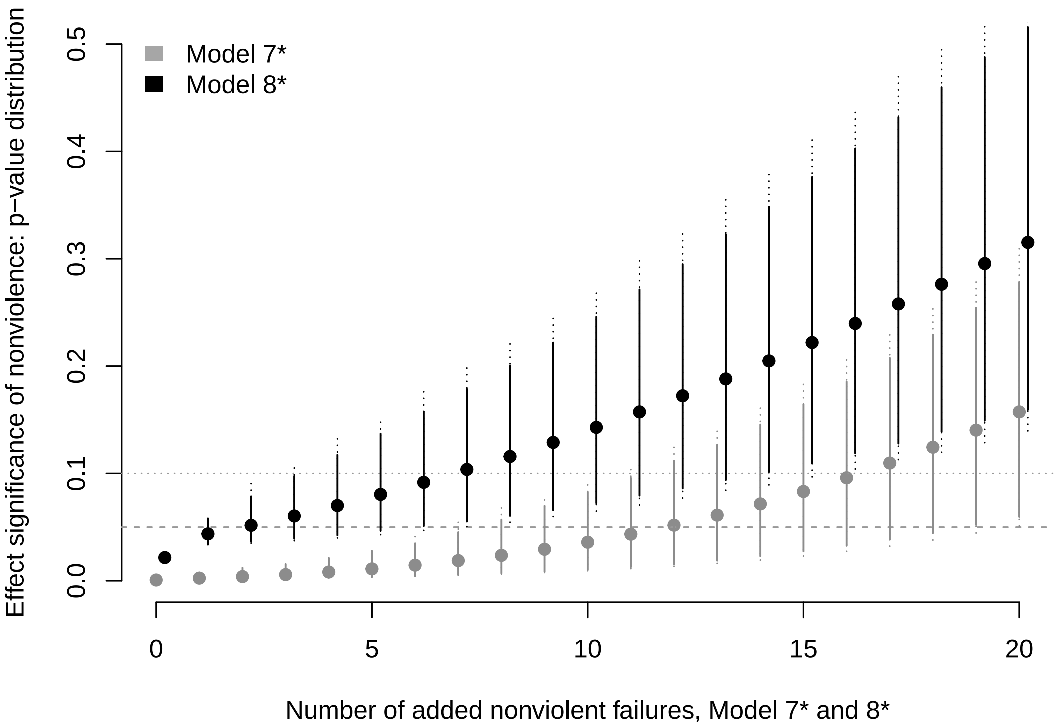
The effect estimate of *nonviolent campaign* with and without (marked with *) campaign size (abbreviated as: *# participants, log*). Models 7 and 8 mirror the results in Table 1 and are included to facilitate comparison.

rule out more alternative explanations. Table A.2 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.

Figures A.2 and A.3 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.2, not conditioning on *number of participants* also increases the effect estimates' robustness to unobserved failed nonviolent campaigns. The simulation approach in Figure A.2, 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.3 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

Figure A.2: Assuming unobserved nonviolent failures, omitting campaign size

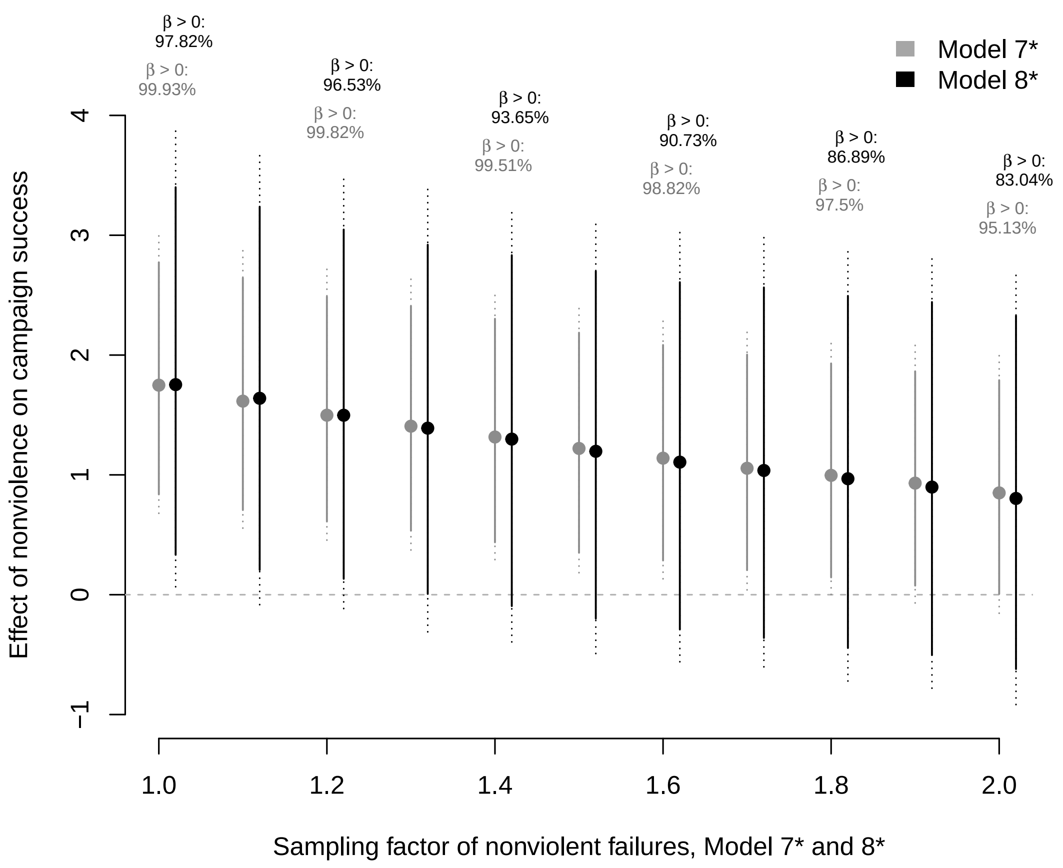


Simulated *p*-value distributions of the effect of *nonviolent campaign* while not conditioning on campaign size, based on Model 7* and Model 8* in Table A.2. Simulation specifications correspond to Figure 3 above.

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 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.2 as linear probability models (LPM). A linear probability model provides an unbiased treatment effect estimate for *nonviolent campaign* while facilitating the sensitivity analysis and en-

Figure A.3: 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.2. Simulation specifications correspond to Figure 6 above.

abling the use of *sensemakr*.[1a] Table A.3 on page A-7 shows the results of the LPMs. Second, I estimate the sensitivity of LPMs 7* and 8*. Table A.4 on page A-7 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 more in the outcome, and vice versa, to lead to the same conclusion. Figure A.4 on page A-8 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

1a. 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.

Table A.3: 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.2.

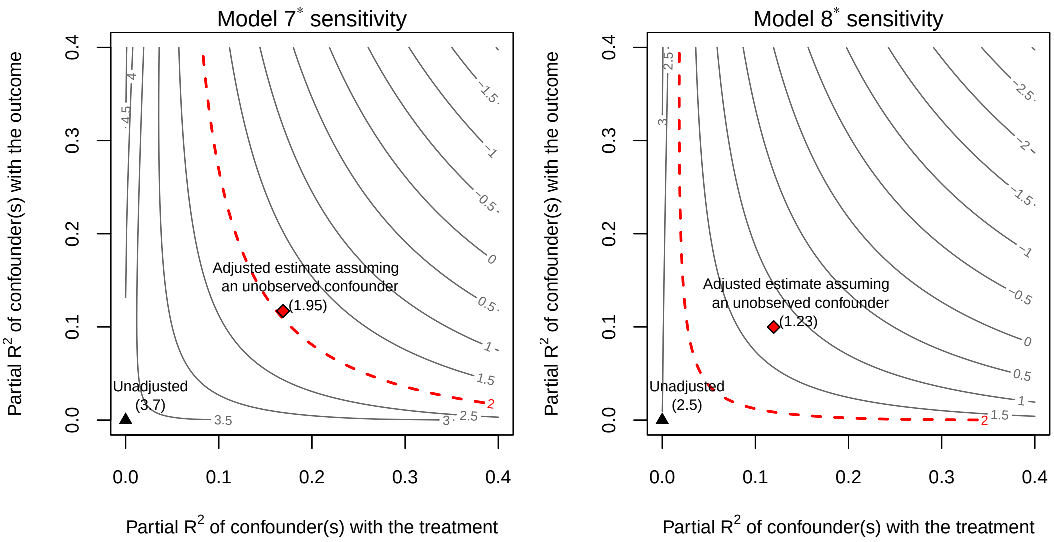
Table A.4: Analysis of sensitivity to omitted variable bias

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.

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.4: Sensitivity analysis when omitting *number of participants*



Visualization based on Table A.4. The hypothetical unobserved confounder is simulated based on *campaigns' goals*. The "unadjusted" marker represents the original coefficient estimate as it appears in Table A.3, 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.4. 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.[2a]

In sum, 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.

2a. 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.