Reporting Methodological Items in Randomized Experiments in Political Science

Isabelle Boutron, Peter John and David J. Torgerson

The ANNALS of the American Academy of Political and Social Science 2010 628: 112
DOI: 10.1177/0002716209351518

Reporting Methodological Items in Randomized Experiments in Political Science

By
ISABELLE BOUTRON,
PETER JOHN,
and
DAVID J. TORGERSON

This article discusses the arguments for using the Consolidated Standards of Reporting Trials (CONSORT) procedures in political science field experiments, with the aim of improving the clarity and transparency of research work and reducing the possibility of bias. The article reviews the background to CONSORT, which is increasingly required for carrying out and reporting trials in healthcare and other disciplines. It sets out the main elements of the scheme and then applies its criteria to evaluate a published Get Out the Vote (GOTV) study by John and Brannan (2008). The CONSORT checklist shows the methods in this article to be clear and transparent but that CONSORT could improve the reporting of turnout experiments, such as details of the numbers going through the trial at each stage. The article argues that applying CONSORT to reports of trials in political science journals is a feasible and desirable objective.

Keywords: trials; field experiments; CONSORT; voter turnout

The randomized controlled trial (RCT) is the best method of preventing selection bias and, in principle, produces unbiased estimates

Isabelle Boutron is a clinical epidemiologist and participated in the development of the extension of the CONSORT Statements for nonpharmacological treatments. *She was supported by a grant from the SFR (Société Française de Rhumatologie) and Lavoisier Program (Ministère des Affaires étrangères et européennes).*

Peter John is the Hallsworth Chair of Governance and is a director of the Institute for Political and Economic Governance (IPEG) in the School of Social Sciences at the University of Manchester, UK. He is author of Analysing Public Policy *(Cassell 1998) and* Local Governance in Western Europe *(Sage 2001).*

David J. Torgerson is the director of the York Trials Unit and is the co-author of Designing Randomised Trials in Health, Education and the Social Sciences *(Palgrave Macmillan 2008).*

NOTE: We thank the attendees of the first Experiments on Government and Politics (EGAP) meeting held at Yale University, April 24-25, 2009, for helpful comments and suggestions on an earlier draft of this article.

DOI: 10.1177/0002716209351518

of an outcome. Only in special circumstances, such as where the selection covariate is clearly known, will estimates using matching and other methods approximate that of an RCT (Shadish, Cook, and Campbell 2002). Randomized trials, however, may be undertaken in a suboptimal fashion. Poorly designed and conducted RCTs may actually be more of a threat to inference than non-randomized controlled trials as the latter are known to be susceptible to selection bias. Consequently their results should be treated more cautiously. The results from an RCT that produces a biased estimate of effect may be accepted uncritically if it is not possible to recognize the difference between a rigorous and a weakly designed RCT. This problem has been recognized in healthcare research, where life-and-death decisions may depend upon the results of a trial. Methodological studies in the 1980s and 1990s found that poorly conducted RCTs generated exaggerated effect sizes compared with the most robustly designed trials (Pocock, Hughes, and Lee 1987; Gore, Jones, and Thompson 1992).

Consequently, a group of trial methodologists and leading medical journal editors formed the Consolidated Standards of Reporting Trials (CONSORT) group (http://www.consort-statement.org/), which produced guidance on the reporting of randomized trials of pharmaceutical products (Altman et al. 2001). These guidelines have been amended to include nonpharmacological interventions (Boutron et al. 2008). CONSORT provides a minimum set of recommendations for reporting RCTs and a standard way for authors to prepare reports of trial findings, which helps full and transparent reporting of the trial. It also is designed to stimulate the critical appraisal and interpretation of experiments. The statement is contained in a 22-item checklist concerning the design, analysis, and interpretation of results, including a flow diagram that shows the progress of all the participants through the trial. The acceptance of these reporting guidelines by journal editors has improved the transparency of published trials. Around three hundred medical journals now require authors to follow CONSORT when reporting a trial. This is not to say a trial has to follow the guidelines in its design, but it has to report whether or not the trial conforms to the CONSORT items. This enables the reader and the systematic reviewer to judge the risk of bias and the applicability of the trial's results. The use of CONSORT has been advocated in the field of educational trials (D. Torgerson and Torgerson 2008), and in this article we argue for its use in political science.

The article is composed of three parts. First, it describes the CONSORT items from the nonpharmacological and cluster trial CONSORT statements and justifies why these are required. Second, it reviews the particular application of RCTs in political science, paying particular attention to voter turnout studies and the kinds of reporting that have been adopted. Third, it takes an example of a trial in political science, John and Brannan's (2008) comparison of door-to-door and telephone Get Out the Vote (GOTV) campaigns in the UK 2005 General Election, and reports the trial using the CONSORT statement to illustrate its utility. The conclusion considers the likely impact of the CONSORT criterion in the political science review process and discusses whether journals in political science should implement CONSORT or a version of it.

Background

RCTs are widely accepted in medical research and in other disciplines, such as education, crime and justice, and other public policy research areas, as the most reliable method to determine the effectiveness of an intervention (Prescott et al. 1999). Other approaches, such as observational studies, can give misleading results (Wood et al. 2008; Kunz and Oxman 1998; Kunz, Vist, and Oxman 2007). In the field of healthcare, several interventions that were deemed to be efficacious on the basis of observational studies turned out to be ineffective or harmful in subsequent RCTs (Abel and Koch 1999). The reason non-randomized studies can be misleading is because of selection bias. Selection bias occurs when participants in a trial are selected into the intervention group on the basis of a variable that is related to outcome. One healthcare example is about the widely accepted view that postmenopausal estrogen replacement therapy reduced cardiovascular disease and strokes (Grady et al. 1992). However, large randomized trials of postmenopausal estrogen replacement therapy showed that this treatment actually increased strokes and heart disease (Writing Group 2002). The previous observational data were misleading because women who took estrogens were either selected to use them by their physicians or approached their physicians to be prescribed estrogens. Such women tended to be different from women who did not use estrogens: they tended to have higher social status, take more exercise, and have a better diet compared with women who did not use the drug. These factors protected the cardiovascular system and misled epidemiologists and clinicians into believing that postmenopausal estrogens could be beneficial for those diseases.

A well-conducted randomized trial ensures that selection bias is eliminated. However, a poorly designed and conducted randomized trial can reintroduce selection bias or produce other biases that may mislead the reader into believing that there is an effect of an intervention when, in truth, there is not. Indeed, a poorly conducted randomized trial may be worse than a well-conducted non-randomized study. Since the latter is acknowledged as being susceptible to selection bias, its results are treated cautiously. For example, a large-cluster randomized trial appeared to show that hip protectors were effective in the prevention of hip fractures (Kannus et al. 2000). But "intention-to-treat" (ITT) analysis was not used. Subsequent, more rigorous trials showed no relationship between hip protectors and lower hip fracture incidence (Birks et al. 2004).

What, then, constitutes a robust randomized trial? There are several key criteria, which we discuss later in this article. However, the most important is transparency of reporting. Any research community that uses RCTs to inform decisions must be able to appraise the internal validity of the trial results—that is, the extent to which systematic errors or bias have been avoided (Clark et al. 1999; Schulz et al. 1995; Guyatt, Sackett, and Cook 1993). Furthermore, a trial should inform wider policy. So for any given trial or systematic review of trials, we need to be able to ascertain whether the results apply outside the setting of the original study, having high external validity.

Unfortunately, this goal has not been achieved in healthcare, mainly because of the inadequate reporting of trials. For example, a systematic review of 519 RCTs published in 2000 highlighted the inadequate reporting of the essential methodological criteria necessary to appraise the internal validity, where such criteria include sample size calculation and the randomization process and handling of attrition (Chan and Altman 2005). Lack of reporting of these details weakens the critical appraisal of results of a trial and makes it difficult to synthesize the research results in systematic reviews and meta-analyses. Healthcare trialists are not the only ones who report methods poorly. Trials undertaken in education, for example, are actually worse when it comes to reporting the details of study design (C. Torgerson et al. 2005).

Experimental studies in political science may also affect policy. As in healthcare, policymakers and other researchers should be in a position to judge whether any randomized trial is of high quality. Many of the same methodological issues relevant to healthcare trials and social science RCTs also will affect trials in political science, so they too need to be reported with clarity. As a first step to improving the reporting of randomized trials in political science, it would seem useful to propose the adoption of some or all of the CONSORT statement for the reporting of such studies. The aim of this article is to describe the CONSORT statement items and convey the rationale for their use. At the same time, it is important to situate the recommendations for political science within the field's general conventions of reporting, for those conventions tend to be more individualistic than in traditional science-based disciplines, for example, by not requiring a structured abstract. Highly structured and diagram-heavy papers may not follow the style of political science journals and might reduce the chances of articles being accepted in a highly competitive environment.

The CONSORT Initiative

Because poor healthcare trials can lead to severe consequences for healthcare policy and could ultimately lead to negative health outcomes (including death), in the 1990s healthcare trial methodologists and journal editors devised a reporting system for RCTs that ensures a minimum quality standard. This initiative led to the CONSORT Statement. Many medical journals have now adopted CONSORT, which means that trial reports should not be published in leading medical journals unless they report their methods in transparent fashion as outlined in the statement.

The CONSORT initiative relies on frequent systematic reviews of all available evidence, on regular meetings governed by consensus, and on continuous assessment of biomedical publications, with regular updates of the guidelines. The most recent update of the CONSORT Statement took place in Montebello (Canada) in January 2007. The dissemination and use of CONSORT guidelines is possible because of the support of a growing number of medical and healthcare journals and editors, including the International Committee of Medical Journal Editors

(ICMJE, the Vancouver Group). Evidence suggests that the use of the CONSORT Statement helps improve the quality of reports of RCTs (Plint et al. 2006). To facilitate the dissemination of the CONSORT Statement, the CONSORT group developed an extension to the statement for abstracts, as well as specific extensions for various trial designs, such as cluster RCTs; for non-inferiority and equivalence trials for various outcomes, such as harm; and for a range of treatments including, recently, nonpharmacological ones. These extensions take into account the specific issues raised in these different contexts.

The CONSORT Statements

The CONSORT checklist recommends the reporting of twenty-two items, as well as a flow diagram, in published articles of RCTs. These items focus on issues considered essential to appraise the risk of bias. We do not detail all the CONSORT items but focus on the essential ones: the randomization process, the blinding of participants and outcome assessors, and the handling of attrition.

Randomization process

In the CONSORT checklist, three items are dedicated to randomization. Random assignment aims to remove the potential of bias in assigning subjects to one intervention or another, which protects against possible systematic connection between the intervention that subjects receive and their prognosis. To achieve this goal, allocation concealment (i.e., a strict implementation of a random allocation sequence) is necessary so that investigators do not know the upcoming assignments. Otherwise, the risk is to not include participants in one intervention arm on the basis of knowledge of their prognosis and investigators' guesses regarding the intervention effect. Evidence suggests that investigators can subvert the allocation concealment process with creative methods (Schulz et al. 1995; Hewitt, Torgerson, and Berger 2009). To avoid such subversion, trials should implement specific methods, such as the use of a secure (independent) third party to do the randomization. Secure allocation is particularly important because empirical investigations (Schulz et al. 1995; Moher et al. 1998) show that, when compared with trials involving adequate concealment, those involving inadequate or unclear allocation concealment yielded up to 40 percent larger estimates of effect. The three items of the CONSORT Statement dedicated to this issue state the need to report (1) the method used to generate the random allocation sequence, (2) the method used to implement the random allocation sequence, and (3) identification of who generated the allocation sequence, enrolled the participants, and assigned them to each group.

Blinding

Blinding, when used in combination with randomization, is essential to limit the occurrence of conscious and unconscious bias. There are several aspects to

blinding. We may wish to blind the participant, whoever delivers the intervention, and the outcome assessor. However, in many sorts of trials such blinding is neither practicable nor possible or even desirable. In pragmatic trials (which measure the effectiveness of a health intervention), it is argued that blinding participants to their intervention is not a good representation of what would happen when the intervention is applied outside the trial. Often, open unblinded trials are more desirable (D. Torgerson and Torgerson 2008). In some trials in political science, such as those that offer an intervention to improve voter turnout (e.g., through canvassing), it is not possible to blind the voter or the canvasser. However, it is very important that the outcome assessor remain blind to group allocation. For example, a voting study would want to ensure that the researcher who is collecting data on voting behavior is blind to the allocation group. Otherwise, researchers may consciously or unconsciously ascertain voting patterns in line with their beliefs rather than with what the data actually show. Methodological studies in healthcare suggest that unblinded outcome assessment is particularly vulnerable to bias. For example, in a multiple sclerosis trial, outcome assessment by an unblinded neurologist revealed an apparent intervention benefit, whereas that by a blinded neurologist did not (Noseworthy et al. 1994). It is unlikely that clinicians are the only ones whose judgments on outcomes may be influenced by their prior beliefs. Such bias may be less of an issue in political science, where results are often observed from verifiable data sources such as electoral registers, but it is still possible, such as in cases where the data are collected within the project.

Blinding is particularly important when measurement of the outcome involves a subjective element. The CONSORT Statement highlights the need to report precisely who was blinded, including details on the method of blinding. In fact, blinding is not well understood. For example, while the terms "single blinding" and "double blinding" are frequently used by researchers and are widely accepted by readers as key markers of validity of an RCT, these terms are not used and interpreted consistently (Devereaux et al. 2001).

Handling of attrition and non-compliance

Violations of the protocol as planned may occur after randomization. Participants may be lost to the follow-up, they may not comply with the allocated intervention, or they may cross over and receive the non-allocated intervention. These protocol violations occur frequently in RCTs and can bias the estimated intervention effect. The recommended strategy is an ITT analysis, an investigation of results of RCTs that compares all participants in the groups to which they were originally randomly assigned (Fergusson et al. 2002; Hollis and Campbell 1999; Schulz et al. 1996). This approach maintains the comparability of intervention groups. The CONSORT Statement recommends reporting a flow diagram of the number of participants randomized, along with the number who complied, withdrew, or were lost to follow-up in each group, as well as the

number analyzed. With the flow diagram, readers should also be able to determine whether all patients were randomized in the group to which they were allocated.

The CONSORT Extension for Nonpharmacological Treatments

Since CONSORT was originally developed for RCTs evaluating drug interventions, it has not always been possible to apply its criteria to the many healthcare trials that are not drug treatments, such as those involving surgery. For example, many non-drug trials cannot use double blinding, and because there may be "therapist" effects, these need to be described in more detail. Consequently CONSORT needed some modification to accommodate these non-drug interventions. Assessing the effectiveness of nonpharmacological interventions, such as in educational medical research, or interventions in other disciplines, such as education and the evaluation of public policies, presents specific issues: the difficulties of blinding, the complexity of interventions, and the possible influence of the skill and expertise of those performing the intervention on the estimates of treatment effects (McCulloch et al. 2002). To ensure that these issues are adequately reported in published RCTs, the CONSORT group developed an extension of the CONSORT Statement for nonpharmacological trials. In February 2006, an international group of thirty individuals, including trialists, methodologists, and journal editors, met in Paris, France. The group reached consensus on specific reporting guidance for RCTs of nonpharmacological interventions (Boutron et al. 2008). Eleven items of the CONSORT checklist were modified. In each case, the modification expanded the text to include a nonpharmacological treatment, and one new item was added on the implementation of the intervention. Below we detail some of the major modifications of the CONSORT checklist.

Complexity of the intervention

Nonpharmacological interventions typically involve several components, each of which can potentially influence the estimated treatment effect (Herbert and Bo 2005; M. Campbell et al. 2000; Hawe, Shiell, and Riley 2004). These interventions are consequently difficult to describe, standardize, and reproduce. The CONSORT extension for nonpharmacological treatment recommends the reporting of all the components of the intervention, as well as additional aspects of how the trial was conducted: the procedure of standardization, the method to assess or enhance treatment adherence, and the details of the intervention as it was actually implemented. These descriptions are necessary to allow for adequate implementation of the treatment into clinical practice. These data are also necessary to facilitate study comparison and inclusion in meta-analyses (Herbert and Bo 2005). Provision of an Internet address for interested readers to access materials the authors used to standardize the interventions could help achieve this goal.

Context influence

In trials assessing nonpharmacological interventions, those providing the interventions are often an integral part of the intervention (Roberts 1999). Consequently, an unequal expertise or skill between two groups could bias treatment effect estimates. Further, the application of an RCT in a different context (lower provider expertise) could produce different results. The CONSORT extension for nonpharmacological treatment insists on this issue and recommends that investigators report (1) eligibility criteria for providers and centers, (2) baseline data for providers, and (3) the reporting of the number of providers or centers performing the intervention in each group and the number of patients treated by each provider or in each center in the flow diagram. These data will improve the understanding of both the internal and external validity of the trial.

Clustering effect

Variation in outcomes is smaller for patients treated by the same care provider (Roberts 1999). Consequently, the assumption that the observed outcomes of participants are independent is false, and observations of participants treated by the same care provider may be clustered. This type of clustering inflates the standard error and reduces the effective sample size, thus reducing the power of the trial (Lee and Thompson 2005). The CONSORT extension for nonpharmacological trials recommends reporting how this issue was handled in the sample size calculation and in the statistical analysis.

Blinding

In non-drug interventions, use of placebo interventions is frequently impossible but is also debated. In fact, the use of placebos has been argued to possibly underestimate the intervention effect (Boutron et al. 2007; D. Torgerson and Torgerson 2008) because placebo interventions may have a specific therapeutic effect linked to the relationship between participants and care providers. Blinding of participants is frequently impossible in nonpharmacological trials, and, consequently, efforts should focus on blinding outcome assessors. Researchers are still working on how best to deal with some of these methodological challenges, and they should report how they handled them to allow progress in understanding these potential biases. This CONSORT extension highlights the need to report these features for all trials of nonpharmacological treatments.

The CONSORT extension for cluster RCTs

Cluster RCTs are often used (and may be the only feasible method) to assess nonpharmacological interventions, particularly because they avoid the threat of contamination of some interventions (such as dietary interventions) if individual randomization is used. Because use of cluster RCTs also raises specific issues, the CONSORT group developed an extension for cluster RCTs (M. J. Campbell,

Elbourne, and Altman 2004). This extension particularly highlights the need to report how the effects of clustering were incorporated into the sample size calculations and how they were incorporated into the analysis. The report should provide a flow diagram showing both the clusters and the progress of individuals through the trial, from assignment to analysis.

The Application of CONSORT to Political Science Trials

Political science has only recently featured randomized controlled trials, which reemerged with voting studies in the 2000s (Green and Gerber 2008). Although there was an experimental tradition in the discipline in the 1930s, it had largely died out with advances in survey research, which seemed at the time to answer most questions in the study of political behavior (Gerber and Green 2003). Partly as a result of the tradition in which political scientists work, they have not been exposed to the conventions of reporting randomized controlled trials, so some procedures, such as giving the power calculations of experiments before their implementation, have not yet been adopted. The CONSORT guidelines could provide a means of catching up with more general reporting standards in the physical sciences and other parts of social science. In the appendix we set out the standard CONSORT items in a table and suggest modifications that can apply to most published trials in political science. We discuss one example here.

John and Brannan (2008) sought to replicate the methods of Gerber and Green (2003) in a field experiment testing the effects of canvassing methods on voter turnout in a single parliamentary seat in the 2005 General Election. The article raises issues of both internal and external validity. We want to know, first, whether we can be confident in its findings and, second, whether these findings are applicable to a wider area than the single geographical location that was the site of the experiment. In Table 1, we apply the CONSORT statement to this particular RCT. We have tried to complete the CONSORT table from data contained within the article. We find there is a good fit and most of the CONSORT items were reported. Some aspects could have been clarified, such as the reason for choosing the sample size and the failure to report tests of the power of the experiment. In addition, one item (item 19) might not apply as adverse events are likely to be very different in politics than in healthcare trials. In healthcare interventions, it is quite common for treatments to have adverse effects; drugs, for instance, may cause gastric side effects, while surgery is prone to infections. For a clinician and patient weighing the merits or hazards of a given treatment these adverse events are very important. However, for political science, this item might be changed to "unexpected events."

In Figure 1, we show the CONSORT flow diagram as applied to the study, which reveals the exact numbers going through the experiment. This diagram allows the reader to understand what is going on with much less effort than just reading the text. (In the original article, the reader would have to calculate the

TABLE 1
Extension of the Consort Statement for Nonpharmacological Interventions—John and Brannan (2008) Study Description

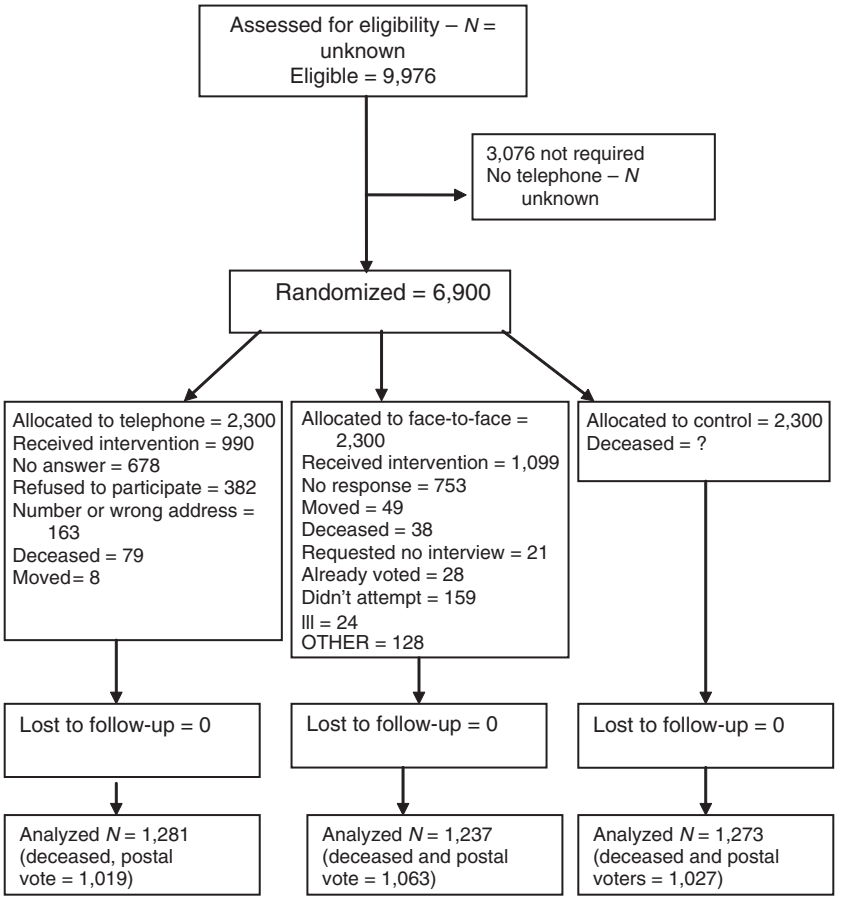
Paper Section and Topic	Item	Standard CONSORT Item; Describe	Comment
Title & Abstract	1	The participants were randomly allocated using a function in the Excel software.	
Introduction			
Background	2	To provide evidence on the effectiveness of canvassing in a UK context.	
Methods			
Participants	3	Participants had to be on the electoral roll and have a landline telephone number.	The results would not be applicable to people who have no public telephone number available.
Interventions	4	Canvassing telephone call or face-to-face visit, preceded by a letter warning of imminent contact. Detailed description of the nonpartisan conversation prompts. Control group received nothing.	
Objectives	5	Can face-to-face or telephone canvassing lead to an increase in the proportion of people who vote in a British General Election?	
Outcomes	6	Main outcome was proportion who voted in the general election. Secondary or process outcomes were the proportion successfully contacted.	
Sample size	7	No prior sample size calculation or justification for sample used.	Note: With 2,300 in each group the trial would have slightly more than 90 percent power to show an absolute 5 percent difference in voting.
Randomization			
Sequence generation	8	Microsoft Excel was used to randomize; no detail was given on stratification—probably specified single random samples of 2,300 from overall sample.	

(continued)

TABLE 1 (continued)

Paper Section and Topic	Item	Standard CONSORT Item; Describe	Comment
Allocation concealment	9	Not clear how concealment was undertaken. The paper did not describe whether the allocation was undertaken by a third party.	
Implementation	10	Not clear.	
Blinding (Masking)	11	Blinding of canvassers not possible or relevant. Does not state whether assessment of official turnout registers was done blindly.	It would have been possible to conceal group allocation from assessment of turnout.
Statistical methods	12	No clear statistical tests used for the ITT analysis; undertook a two-stage regression for instrumental variable analysis.	
Results			
Participant flow	13	Detailed description given in Tables 1 & 2 about reasons for non-contact of participants. Not possible to estimate total initial sample before exclusions due to lack of telephone landline.	
Recruitment	14	Not specified.	
Baseline data	15	Not possible as electoral roll gives limited demographic detail of electors.	
Numbers analyzed	16	In main table of results does not give both numerator and denominator.	
Outcomes and estimation	17	Provision of standard error but not confidence intervals.	
Ancillary analyses	18	None performed.	
Adverse events	19	None reported.	Adverse events may not be relevant here.
Discussion			
Interpretation	20	Interpretation draws on earlier U.S. literature and shows similar findings.	
Generalizability	21	May have poor generalizability as it was a single safe constituency in a relatively poor area and may not apply to wealthier areas.	
Overall evidence	22	Draws on past evidence that shows a marginal impact.	

FIGURE 1
Consort Flow Chart for John and Brannan (2008)



difference between the randomized sample of 2,300 and the analyzed numbers in the tables to work out the numbers of deceased and postal voters the researchers removed.) The clarity provided by the diagram would thus have benefited the article.

Discussion and Conclusion

The discussion of the background to the CONSORT guidelines and their implementation is designed to show the importance of the reporting of randomized

controlled trials because of the dangers to scientific understanding and inference from poorly reported trials. If the people using trials come to false conclusions, particularly if they conclude there is an effect when there might not be one, then the whole point of doing RCTs—to provide valid and robust knowledge from which to make policy or other decisions—is undermined. The trustworthiness of RCTs is obviously crucial in healthcare, but it is also important in policy-relevant areas such as voter turnout. In addition, the guidelines act as an extra discipline in the research process. They encourage researchers to adopt the highest standards in the design of their research, spurring them to address issues of validity and reliability before they complete their projects. To this end, full transparency in reporting a trial means the reader can be assured that the science behind a study is at the highest standard. The fellow researcher can trust the study's inferences or fairly assess its limitations when designing replications or extensions. If this argument is accepted, then political scientists should consider adopting the CONSORT criteria in carrying out and reporting experiments.

We presented a case study of John and Brannan (2008) to illustrate the advantages of CONSORT. Because of the large number of voting studies that use the same method and form of reporting (see Green and Gerber 2008), we think it is a fair assessment of the state of methods and reporting in the field. The CONSORT checklist and flow chart would look similar in most of these studies, though without the complexities caused by the UK electoral registration system that affected the sample size in the John and Brannan case. There is no doubt that the CONSORT reporting is cleaner and would have made the article and those like it more explicit in its account of the stages of the trial. It would help the reader better understand the study, in particular the numbers of subjects at each stage. It would also have been useful to see the calculations of effect size and power before the experiment. So in that sense, CONSORT offers an advantage for researchers and that alone could be a reason for its adoption.

The bigger question is whether the CONSORT guidelines would have produced better experiments in political science. Here, the presentation of the data shows that the experiment was done properly, and it reflects the high standards of the reporting of methodological issues in political science generally. It would not have taken much effort to have the article report the CONSORT checklist. In addition, political science experiments that rely on publicly validated data or that are done by independent survey companies may not have the same vulnerability to violations of the experimental design as other disciplines that have more direct contact with their research subjects. It partly reflects the difficulty of doing research on politicians, political actors, and the citizens themselves that the unit of measurement tends not to be based on direct observations of those actors, though experiments on political actors are getting more common.

CONSORT should not inhibit the carrying out and publication of experiments. There may be justifiable reasons that researchers may not be able to apply all items of the statement. For example, while in an ideal world we would identify an important difference that we wish to detect between groups and plan our study accordingly, in the real world, we may not be able to do so. The sample size

might be governed by resource or time availability. Or a sample size might have been fixed in advance through policy or political constraints. Nevertheless, it is thought to be good practice to report the underlying reasons for the sample size, whether they are statistical or practical in nature. Transparency is the key, so that readers can make their own judgments.

A generally strong methodological tradition and an often favorable research environment in political science are not reasons for complacency, particularly as experiments diffuse in the discipline. A checklist provides discipline, helping the researcher carry out a study efficiently. It is possible that future experiments will have more direct contact with research subjects, especially as experimental research expands out from voter turnout studies. At the same time, there is a move for more transparency in the reporting of political science methods more generally that the CONSORT initiative neatly complements. At the moment, the bulky character of the CONSORT reporting requirements might not be quite the norm for journals used to more economical forms of presentation (though they could perhaps be adopted in an online appendix). They could conceivably even put off reviewers and journal editors, perhaps subjecting trials to a higher level of scrutiny than papers using other methods. But greater detail about the methods, such as that provided by CONSORT, is probably going to be more of a norm across the whole of political science. And there is a final advantage: these guidelines will help experimental researchers outside political science to understand political science experiments.

Appendix

Modified Extension of the CONSORT Statement for Political Science Trials Adapted from the Extension for Nonpharmacological Treatments

Paper Section and Topic	Item	Standard CONSORT Item; Describe	Extension for Reports of Trials in Political Science; in Addition
Title & Abstract	1	How participants were allocated to interventions (e.g., "random allocation," "randomized," or "randomly assigned")	In the abstract, description of the experimental intervention, comparator, intervention providers, centers, and blinding status
Introduction			
Background	2	Scientific background and explanation of rationale	
Methods			
Participants	3	Eligibility criteria for participants and the settings and locations where the data were collected	When applicable, eligibility criteria for centers and those performing the interventions
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered	Precise details of both the experimental intervention and comparator
	4.A		Description of the different components of the interventions and, when applicable, descriptions of the procedure for tailoring the interventions to individual participants
	4.B		Details of how the interventions were standardized
	4.C		Details of how adherence of intervention providers with the protocol was assessed or enhanced
Objectives	5	Specific objectives and hypotheses	
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the	

(continued)

Appendix (continued)

Paper Section and Topic	Item	Standard CONSORT Item; Describe	Extension for Reports of Trials in Political Science; in Addition
		quality of measurements (e.g., multiple observations, training of assessors)	
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules	When applicable, details of whether and how the clustering by intervention providers or centers were addressed
Randomization			
Sequence generation	8	Method used to generate the random allocation sequence, including details of any restriction (e.g., blocking, stratification)	When applicable, how intervention providers were allocated to each trial group
Allocation concealment	9	Method used to implement the random allocation sequence (e.g., numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned	
Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups	
Blinding (Masking)	11.A	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment	Whether or not those administering co-interventions were blinded to group assignment
	11.B		If blinded, method of blinding and description of the similarity of interventions[a]
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s). Methods for additional analyses, such as subgroup analyses and adjusted analyses	When applicable, details of whether and how the clustering by intervention providers or centres was addressed

(continued)

Appendix (continued)

Paper Section and Topic	Item	Standard CONSORT Item; Describe	Extension for Reports of Trials in Political Science; in Addition
Results			
Participant flow	13	Flow of participants through each stage (a diagram is strongly recommended). Specifically, for each group, report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analyzed for the primary outcome. Describe protocol deviations from study as planned, together with reasons	The number of intervention providers or centers performing the intervention in each group and the number of participants treated by each intervention provider or in each center
Implementation of intervention	New item		Details of the experimental intervention and comparator as they were implemented
Recruitment	14	Dates defining the periods of recruitment and follow-up	
Baseline data	15	Baseline demographic and clinical characteristics of each group	Baseline characteristics of each group and when applicable, a description of intervention providers (case volume, qualification, expertise, etc.) and center (volume) in each group
Numbers analyzed	16	Number of participants (denominator) in each group included in each analysis and whether analysis was by ITT; state the results in absolute numbers when feasible (e.g., 10/20, not 50 percent)	
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group and the estimated effect size and its precision (e.g., 95 percent confidence interval)	

(continued)

Appendix (continued)

Paper Section and Topic	Item	Standard CONSORT Item; Describe	Extension for Reports of Trials in Political Science; in Addition
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those pre-specified and those exploratory	
Unexpected events (in CONSORT, adverse events)	19	All important adverse events or side effects in each intervention group	All important unexpected events (adverse events or side effects) in each intervention group
Discussion			
Interpretation	20	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision, and the dangers associated with multiplicity of analyses and outcomes	Additionally take into account the choice of the comparator, lack of or partial blinding, and unequal expertise of intervention providers or centers in each group
Generalizability	21	Generalizability (external validity) of the trial findings	Generalizability (external validity) of the trial findings according to the intervention, comparators, participants, intervention providers, and centers involved in the trial
Overall evidence	22	General interpretation of the results in the context of current evidence	

a. This item was modified in the 2007 revised version of the CONSORT checklist.

References

Abel, U., and A. Koch. 1999. The role of randomization in clinical studies: Myths and beliefs. *Journal of Clinical Epidemiology* 52 (6): 487-97.

Altman, Douglas G., Kenneth F. Schulz, David Moher, Matthias Egger, Frank Davidoff, Diana Elbourne, Peter C. Gøtzsche, and Thomas Lang. 2001. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Annals of Internal Medicine* 134 (8): 663-94.

Birks, Yvonne F., Jill Porthouse, Caroline Addie, Karen Loughney, Lucy Saxon, Mike Baverstock, Roger M. Francis, David M. Reid, Ian Watt, and David J. Torgerson. 2004. Randomized controlled trial of hip protectors among women living in the community. *Osteoporosis International* 15:701-6.

Boutron Isabelle, Lydia Guittet, Candice Estellat, David Moher, Asbjørn, Hróbjartsson, and Philippe Ravaud. 2007. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Medicine* 4 (2): e61.

Boutron, Isabelle, David Moher, Douglas G. Altman, Kenneth F. Schulz, and Philippe Ravaud. 2008. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: Explanation and elaboration. *Annals of Internal Medicine* 148 (4): 295-309.

Campbell, Marion J., Diana R. Elbourne, and Douglas G. Altman. 2004. Extending CONSORT to include cluster trials. *BMJ* 328 (7441): 654-55.

Campbell, Michelle, Ray Fitzpatrick, Andrew Haines, Ann L. Kinmonth, Peter Sandercock, David Spiegelhalter, and Peter Tyrer. 2000. Framework for design and evaluation of complex interventions to improve health. *BMJ* 321 (7262): 694-96.

Chan, AnWen, and Douglas G. Altman. 2005. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 365 (9465): 1159-62.

Clark, H. D., G. A. Wells, C. Huet, F. A. McAlister, L. R. Salmi, D. Fergusson, and A. Laupacis. 1999. Assessing the quality of randomized trials: Reliability of the Jadad scale. *Controlled Clinical Trials Journal* 20 (5): 448-52.

Devereaux, P. J., Braden J. Manns, William A. Ghali, Hude Quan, Christina Lacchetti, Victor M. Montori, Mohit Bhandari, and Gordon H. Guyatt. 2001. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 285 (15): 2000-2003.

Fergusson, Dean, Shawn D. Aaron, Gordon Guyatt, and Paul Hérbert. 2002. Post-randomisation exclusions: The intention to treat principle and excluding patients from analysis. *BMJ* 325 (7365): 652-54.

Gerber, Alan S., and Donald P. Green. 2003. Reclaiming the experimental tradition in political science. In *Political science: The state of discipline*, ed. Helen V. Milner and Ira Katznelson, 805-32. New York: Norton.

Gore, S. M., G. Jones, and S. G. Thompson. 1992. The Lancet's statistical review process: Areas for improvement by authors. *The Lancet* 340:100-102.

Grady, D., S. B. Reuben, D. B. Pettiti, C. S. Fox, D. Black, B. Ettinger, V. L. Ernster, and S. R. Cummings. 1992. Hormone therapy to prevent disease and prolong life in postmenopausal women. *Annals of Internal Medicine* 117:1016-37.

Green, Donald P., and Alan Gerber. 2008. *Get out the vote*. Washington, DC: Brookings Institution.

Guyatt, Gordon H., David L. Sackett, and Deborah J. Cook. 1993. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 270 (21): 2598-2601.

Hawe, Penelope, Alan Shiell, and Therese Riley. 2004. Complex interventions: How "out of control" can a randomised controlled trial be? *BMJ* 328 (7455): 1561-63.

Herbert, Robert D., and Kari Bo. 2005. Analysis of quality of interventions in systematic reviews. *BMJ* 331 (7515): 507-9.

Hewitt, Catherine E., David J. Torgerson, and Vance W. Berger. 2009. Potential for technical errors and subverted allocation can be reduced if certain guidelines are followed: Examples from a Web-based survey. *Journal of Clinical Epidemiology* 62:261-69.

Hollis, Sally, and Fiona Campbell. 1999. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 319 (7211): 670-74.

John, Peter, and Tessa Brannan. 2008. How different are telephoning and canvassing? Results from a Get Out the Vote field experiment in the UK 2005 General Election. *British Journal of Political Science* 38:565-74.

Kannus, Pekka, Jari Parkkari, Seppo Niemi, Matti Pasanen, Mika Palvanen, Markku Järvinen, and Ilkka Vuori. 2000. Prevention of hip fracture in elderly people with use of a hip protector. *New England Journal of Medicine* 343:1506-13.

Kunz, Regina, and Andrew D. Oxman. 1998. The unpredictability paradox: Review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 317 (7167): 1185-90.

Kunz, Regina, Gunn E. Vist, and Andrew D. Oxman. 2007. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database of Systematic Reviews* 2:MR000012.

Lee, Katherine J., and Simon G. Thompson. 2005. Clustering by health professional in individually randomised trials. *BMJ* 330 (7483): 142-44.

McCulloch, Peter, Irving Taylor, Mitsuru Sasako, Bryony Lovett, and Damian Griffin. 2002. Randomised trials in surgery: Problems and possible solutions. *BMJ* 324 (7351): 1448-51.

Moher, David, Ba' Pham, Alison Jones, Deborah J. Cook, Alejandro R. Jadad, Michael Moher, Peter Tugwell, and Terry P. Klassen. 1998. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *The Lancet* 352 (9128): 609-13.

Noseworthy, J., G. C. Ebers, M. K. Vandervoot, R. E. Farquhar, E. Yetisir, and R. Roberts. 1994. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology* 44 (1): 16-20.

Plint, Amy C., David Moher, Andra Morrison, Kenneth Schulz, Douglas G. Altman, Catherine Hill, and Isabelle Gaboury. 2006. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Medical Journal of Australia* 185 (5): 263-67.

Pocock, Stuart J., M. D. Hughes, and R. J. Lee. 1987. Statistical problems in the reporting of clinical trials. *New England Journal of Medicine* 317:426-32.

Prescott, R. J., C. E. Counsell, William J. Gillespie, Adrian M. Grant, Ian T. Russell, S. Kiauka, I. R. Colthart, Susan Ross, S. M. Shepherd, and Daphne Russell. 1999. Factors that limit the quality, number and progress of randomised controlled trials. *Health Technology Assessment* 3 (20): 1-143.

Roberts, Chris. 1999. The implications of variation in outcome between health professionals for the design and analysis of randomized controlled trials. *Statistics in Medicine* 18 (19): 2605-15.

Schulz, Kenneth F., Iain Chalmers, Richard J. Hayes, and Douglas G. Altman. 1995. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 273 (5): 408-12.

Schulz, Kenneth F., David A. Grimes, Douglas G. Altman, and Richard J. Hayes. 1996. Blinding and exclusions after allocation in randomised controlled trials: Survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 312 (7033): 742-44.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Torgerson, Carole J., David J. Torgerson, Yvonne F. Birks, and Jill Porthouse. 2005. A comparison of randomised controlled trials in health and education. *British Educational Research Journal* 31:761-85.

Torgerson, David J., and Carole J. Torgerson. 2008. *Designing randomised trials in health, education and the social sciences*. Houndmills, UK: Palgrave Macmillan.

Wood, Lesley, Mathias Egger, Lise L. Gluud, Kenneth F. Schulz, Peter Jüni, Douglas G. Altman, Christian Gluud, Richard M. Martin, Anthony J. G. Wood, and Jonathan A. C. Sterne. 2008. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: Meta-epidemiological study. *BMJ* 336 (7644): 601-5.

Writing Group for the Women's Health Initiative Investigators. 2002. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the women's health initiative randomised controlled trial. *JAMA* 288:321-33.