Reporting methodological items in randomised experiments in political science

Isabelle Boutron (Centre for Statistics in Medicine University of Oxford, OX2 6UD, isabelle.boutron@bch.aphp.fr)

Peter John (School of Social Sciences, University of Manchester, Oxford Road, Manchester M13 9PL UK, peter.john@manchester.ac.uk)

David J. Torgerson (Trials Unit, Area 4, Seebohm Rowntree Building, Dept of Health Sciences, University of York, York YO10 5DD UK).

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 (Societe Francaise 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* and *Local Governance in Western Europe* (2001).

David Torgerson is the Director of the York Trials Unit and is the author of *Designing Randomised Trials in Health, Education and the Social Sciences* (2008).

Article prepared for a special issue of the *Annals of Political and Social Sciences*

Abstract

This article discusses the arguments for using the CONSORT procedures in political science field experiments, with 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 required for carrying out and reporting trials in health 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 study, 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 details of the numbers going through the trial at each stage. The article argues that applying CONSORT to all trials in political science trials is a feasible and desirable objective.

Key words: trials, field experiments, CONSORT, voter turnout

Introduction

The randomized controlled trial (RCT) is the best method of preventing selection bias and, in principle, produces unbiased estimates of an outcome. Only in special circumstances, such as where the selection covariate is clearly known, will estimates using matching and other methods approximate to that of a RCT (Shadish et al 2008). 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 and consequently their results should be treated more cautiously. The results from a 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 et al, 1987; Gore et al, 1992).

Consequently, a group of trial methodologists and leading medical journal editors formed the CONSORT group (http://www.consort-statement.org/), which produced guidance on the reporting of randomised trials of pharmaceutical products (Altman et al 2001). These guidelines have been amended to include non-pharmacological 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 the transparent

reporting of the trial. It also is designed to stimulate the critical appraisal and interpretation of experiments. The statement is contained in a twenty-two-item checklist concerning the design, analysis and interpretation of results, with a flow diagram that shows the progress of all the participants through the trial. The implementation of these reporting guidelines by editors has improved the transparency of published trials. Around 300 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 (Torgerson and Torgerson, 2005) and in this article we argue for its use in political science.

The article is in three parts. First, it describes the CONSORT items from the non-pharmacological and cluster trial CONSORT statements and justify 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 the political sciences and report this using the CONSORT statement to illustrate its utility: John and Brannan's (2008) comparison of a door-to-door and telephone Get Out the Vote study in the UK 2005 General Election. The conclusion considers the likely impact of the CONSORT criterion in the political science review process and discusses whether papers in political science should implement CONSORT or a version of it.

Background

Randomized controlled trials (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 et al., 1998; Kunz et al. 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 et al, 1999). The reason that non-randomised 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 health care example is about the widely accepted view that post-menopausal oestrogen replacement therapy reduced cardiovascular disease and strokes (Grady et al, 1992). However, large randomized trials of post-menopausal oestrogen replacement therapy showed this treatment actually increased strokes and heart disease (Writing Group, 2002). The previous observational data were misleading because women who took oestrogens were either selected to use them by their physicians or approached their physicians to be prescribed oestrogens. Such women tended to be different from women who did not use oestrogens: they tended to have higher social status, to take more exercise and to 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 post-menopausal oestrogens 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 that a well-conducted non-randomised study. As the latter is acknowledged as being susceptible to selection bias its results are then 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 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 (Clark et al, 1999; Schulz et al, 1995; Guyatt et al, 1993) (i.e., the extent to which systematic errors or bias has been avoided). 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 health care, 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, such as sample size calculation, the randomisation 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 synthesise of the research results in systematic reviews and meta-analyses. Health care 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 (Torgerson et al, 2005).

Experimental studies in political sciences may also affect policy. Policy makers and other researchers should, like in health care, be in a position to judge whether any randomised trial is of high quality. Many of the same methodological issues relevant to health care trials and social science RCTs also will affect trials in the political science so they too need to be reported with clarity. As a first step to improving the reporting of randomised trials in the political sciences 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 rationale for their use. At the same time it is important to situate the recommendations for political science within its general conventions of reporting, which tends to be more individualistic than in traditional science-based disciplines, for example by not having a structured abstract. Highly structured and diagram-heavy papers may not follow the style of political science journals and might reduce the chance of articles being accepted in a highly competitive environment.

The CONSORT initiative

Because poor health care trials can lead to severe consequences for health policy and could ultimately lead to negative health outcomes (including death), in the 1990s health care trial methodologists and journal editors devised a reporting system for RCTs that ensures a minimum quality standard. This initiative led to the Consolidated Standards of Reporting Trials (CONSORT) Statement (http://www.consort-statement.org/). Many medical journals have now adopted this, 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, regular meetings governed by consensus and 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 these guidelines is possible because of the support of a growing number of medical and health care 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, non-inferiority and equivalence trials; for various outcomes, such as harm; and for a range of treatments including, recently, non-pharmacological 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 will 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, 1995; Hewitt et al, 2009). To avoid such subversion, trials should implement specific methods, such as the use of secure (independent), third party to do the randomisation. 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 forty per cent 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) 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, whosoever delivers the intervention and the outcome assessor. However, in many sorts of trials this 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 was applied outside the trial. Often open unblinded trials are more desirable (Torgerson and Torgerson, 2008). Some trials in political science, which offer an intervention to improve voter turnout (e.g., through canvassing), cannot blind the voter and is it not

possible to blind the canvasser either. However, it is very important that the outcome assessor remains blind to group allocation. For example, a voting study would want to ensure that the researcher who is collecting data on voting behaviour is blind to the allocation group. Otherwise the researcher may consciously or unconsciously ascertain voting patterns in line with her or his beliefs rather than what the data actually show. Methodological studies in health care 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 judgements on outcomes may be influenced by their prior beliefs! This is probably less of an issue in political science where results are often observed from verifiable data sources like electoral registers. But it cannot be guaranteed, such as 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, with details on the method of blinding. In fact, blinding is not well understood. For example, the terms 'single blinding' and 'double blinding' are frequently used by researchers and are widely accepted by readers as key markers of validity of a RCT. But these terms are not used and interpreted consistently (Devereaux et al, 2001).

Handling of attrition and non-compliance

Violations to 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 intention-to treat analysis, an analysis of results of RCTs that compares all participants in the groups to which they were originally randomly assigned (Ferguson et al, 2002; Hollis and Campbell, 1999; Schulz et al, 1996). This approach maintains the comparability of intervention groups. The CONSORT Statement recommend the reporting of a flow diagram with the number of participants randomized, the number who complied, withdrew and 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 non-pharmacological treatments

Although CONSORT was originally developed for RCTs evaluating drug interventions, it is clear that many health care trials are not drug treatments, such as surgery, and consequently CONSORT did not quite fit these non-pharmaceutical contexts. 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 non-pharmacologic 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 these issues are inadequately reported in published RCTs, the CONSORT group developed an extension of the CONSORT Statement for non-pharmacological trials. In February 2006, an international group of thirty individuals, including trialists, methodologists and journal editors, met in Paris. The group reached consensus on specific reporting guidance for RCTs of non-pharmacological interventions (Boutron et al, 2008). Eleven items of the CONSORT checklist were modified. In each case, the modification expanded the text to include a non-pharmacological 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

Non-pharmacological interventions typically involve several components, each of which can potentially influence the estimated treatment effect (Herbet and Bo, 2005; Campbell et al, 2000; Hawe et al, 2004). These interventions are consequently difficult to describe, standardize and reproduce. The CONSORT extension for non-pharmacological 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 (Herbet 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 non-pharmacological 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 non-pharmacological treatment insists on this issue and recommends that investigators report: 1) eligibility criteria for providers and centres; 2) baseline data for providers; and 3) the reporting of the number of providers or centres performing the intervention in each group and the number of patients treated by each provider or in each centre 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 (Lee and Thompson, 2005). 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 non-pharmacological 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; 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 non-pharmacologic 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 non-pharmacologic treatments.

The CONSORT extension for cluster RCTs

Cluster RCTs are often used to assess non-pharmacologic interventions, particularly because they avoid the threat of contamination of some interventions (such as dietary

interventions) if individual randomisation is used and may be the only feasible method. Because use of cluster RCTs also raises specific issues, the CONSORT group developed an extension for cluster RCTs (Campbell et al, 2004). This extension particularly highlights the need to report how the effects of clustering were incorporated into the sample size calculations and how were incorporated into the analysis. The report should provide in the flow diagram showing the 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 randomised controlled trials, which emerged with voting studies in the 2000s (Gerber and Green 2008). Though there was an experimental tradition 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 behaviour (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 randomised controlled trials, so some procedures, such as giving the power calculations of experiments before their implementation, have not been yet adopted. The CONSORT guidelines could provide a means of catching up with more general reporting standards in science and other parts of social science. We discuss one example here.

John and Brannan (2008) sought to replicate the methods of Gerber and Green (2000) in a field experiment testing the effects of difference 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 whether we be confident in its findings and secondly are these findings applicable to a wider area than the single geographical area that was the site of the experiment. In Table 1 we apply the CONSORT statement to this particular RCT. In the table 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 health trials. In health care interventions it is quite common for treatments to have adverse effects. Drugs for instance may cause gastric side-effects, whilst surgery in prone to infections. For a clinician and patient weighing up the merits or hazards of a given treatment these adverse events are very important. This 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 allows the reader to understand what is going on with much less effort than just reading the text. Currently, the reader has to calculate the difference between the randomised sample of 2,300 and the analysed numbers in the tables of the article to work out the

numbers of deceased and postal voters the researchers removed. This clarity would have benefited the article.

Discussion and Conclusions

The discussion of the background to the CONSORT guidelines and their implementation is designed to show how serious is the reporting of randomised controlled trials. This is because of the dangers to scientific understanding and inference from poorly reported trials. If the people using trials come to false conclusions, in particular concluding there is an effect when there might not be or 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. This is obviously crucial in medicine and health where people's health is at risk, but it is important in policy relevant areas such as voter turnout. In addition, the guidelines acts as an extra discipline in the research process, affecting how researchers do the research if they know there are very transparent means of reporting it. It encourages them to have the highest standards in the design of the research which can help them address issues of validity and reliability before they complete their projects. To this end, fully transparent standards of reporting a trial a means the reader can ensure the science behind a study is done 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 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 impacted on 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 understand the study more, in particular the numbers of subjects at each stage. It would have been useful to see the calculations of effect size and power before the experiment. So in that sense CONSORT offers an advantage for researcher and could alone 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 show 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 it report the CONSORT checklist. In addition, political science experiments that rely on publicly validated data or that done by independent survey companies that 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 is 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, whilst 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 this. 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 judgements.

A generally strong methodological tradition and an often favourable research environment in political science are not reasons for complacency, particularly as experiments diffuse in the discipline. A checklist is a good discipline to have, helping the researcher carry out a study efficiently, much in same the way checklists assist other activities, such as the citizen filling in a tax return or a surgeon carrying out an operation in the proper sequence. It is possible that future experiments will have more direct contact with the research subjects, especially as experimental research expands out from voter turnout studies. And at the same time there is a move for more transparency in reporting of political science methods more generally which the CONSORT initiative neatly complements. At the moment, the bulky character of the CONSORT reporting requirements might not be quite the current 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 an higher level of scrutiny than papers using other methods so creating a non-level playing field. But greater detail about the methods, like CONSORT, is probably going to be more of a norm across the whole of political science. And there is one final advantage we have not mentioned: these guidelines will allow experimental researchers outside political science to understand political science experiments.

Acknowledgements

We thank the attendees of the first Experiments on Government and Politics (GAP) meeting held at Yale University April 24-25 for helpful comments and suggestions on an earlier draft of this manuscript.

References

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

Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med. 2001;134(8):663-94

Birks YF, Porthouse J, Addie C, Loughney K, Saxon L, Baverstock M, et al. (2004) Randomized controlled trial of hip protectors among women living in the community. Osteoporosis International 15, 701-706.

Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med. 2008;148(4):295-309.

Boutron I, Guittet L, Estellat C, Moher D, Hrobjartsson A, Ravaud P. Reporting Methods of Blinding in Randomized Trials Assessing Nonpharmacological Treatments. PLoS Med. 2007;4(2):e61.

Campbell M, Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health. Bmj. 2000;321(7262):694-6.

Campbell MJ. Extending CONSORT to include cluster trials. Bmj. 2004;328(7441):654-5.

Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. Lancet. 2005;365(9465):1159-62

Clark HD, Wells GA, Huet C, et al. Assessing the quality of randomized trials: reliability of the Jadad scale. Control Clin Trials. 1999;20(5):448-52.

Devereaux PJ, Manns BJ, Ghali WA, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. Jama. 2001;285(15):2000-3.

Fergusson D, Aaron SD, Guyatt G, Hebert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. Bmj. 2002;325(7365):652-4.

Gerber, A. and Green, D. (2003), 'Reclaiming the Experimental Tradition in political science', Special issue on experimental methods in *Political Analysis*, Vol 10, No.

Gore SM, Jones G, Thompson SG. (1992) The Lancet's statistical review process: areas for improvement by authors. Lancet 340, 100-2.

Grady D, Rueben SB, Pettiti DB, et al. (1992) Hormone therapy to prevent disease and prolong life in postmenopausal women. Annals of Internal Medicine 117, 1016-1037.

Green, D. and Gerber, A. (2008), *Get Out The Vote* Washington: Brookings

Guyatt GH, Sackett DL, Cook DJ. 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. 1993;270(21):2598-601.

Hawe P, Shiell A, Riley T. Complex interventions: how "out of control" can a randomised controlled trial be? Bmj. 2004;328(7455):1561-3.

Hewitt CE, Torgerson DJ, Berger V. 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 2009;62;261-269

Herbert RD, Bo K. Analysis of quality of interventions in systematic reviews. Bmj. 2005;331(7515):507-9.

Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. Bmj. 1999;319(7211):670-4.

John, P. and Brannan, T. 2008. How different are telephoning and canvassing? A Get Out The Vote field experiment in the UK 2005 General Election, *British Journal of Political Science,* 38: 565-574.

Kannus P, Parkkari J, Niemi S, Pasanen M, Palvanen M, Järvinen M, Vuori I. (2000) Prevention of hip fracture in elderly people with use of a hip protector. New England Journal of Medicine 343, 1506-1513.

Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. Bmj. 1998;317(7167):1185-90.

Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev. 2007(2):MR000012.

Lee KJ, Thompson SG. Clustering by health professional in individually randomised trials. Bmj. 2005;330(7483):142-4.

McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. Bmj. 2002;324(7351):1448-51.

Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet. 1998;352(9128):609-13

Noseworthy J, Ebers GC, Vandervoot MK, Farquhar RE, Yetisir E, Roberts R. (1994) The impact of blinding on the results of a randomized, placebo- controlled multiple sclerosis clinical trial. Neurology 44, 16-20.

Plint AC, Moher D, Morrison A, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Med J Aust. 2006;185(5):263-7

Pocock SJ, Hughes MD, Lee RJ. (1987) Statistical problems in the reporting of clinical trials. New England Journal of Medicine 317, 426-32

Prescott RJ, Counsell CE, Gillespie WJ, et al. Factors that limit the quality, number and progress of randomised controlled trials. Health Technol Assess. 1999;3(20):1-143.

Roberts C. The implications of variation in outcome between health professionals for the design and analysis of randomized controlled trials. Stat Med. 1999;18(19):2605-15.

Schulz KF. Subverting randomization in controlled trials. Jama. 1995;274(18):1456-8.

Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. Bmj. 1996;312(7033):742-4.

Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. Jama. 1995;273(5):408-12.

Torgerson DJ, Torgerson CJ. Designing randomised trials in health, education and the social sciences. 2009, Palgrave Macmillan.

Torgerson CJ, Torgerson DJ, Birks YF, Porthouse J. (2005) A comparison of randomised controlled trials in health and education. British Educational Research Journal 31,761-785.

Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. Bmj. 2008;336(7644):601-5.

Writing Group for the Women's Health Initiative Investigators. (2002) Risks and benefits of oestrogen plus progestin in healthy postmenopausal women: Principal results from the women's health initiative randomised controlled trial. Journal of the American Medical Association 288, 321-33

Table 1: Extension of the CONSORT Statement for nonpharmacological interventions – John and Brannan (2008) study description

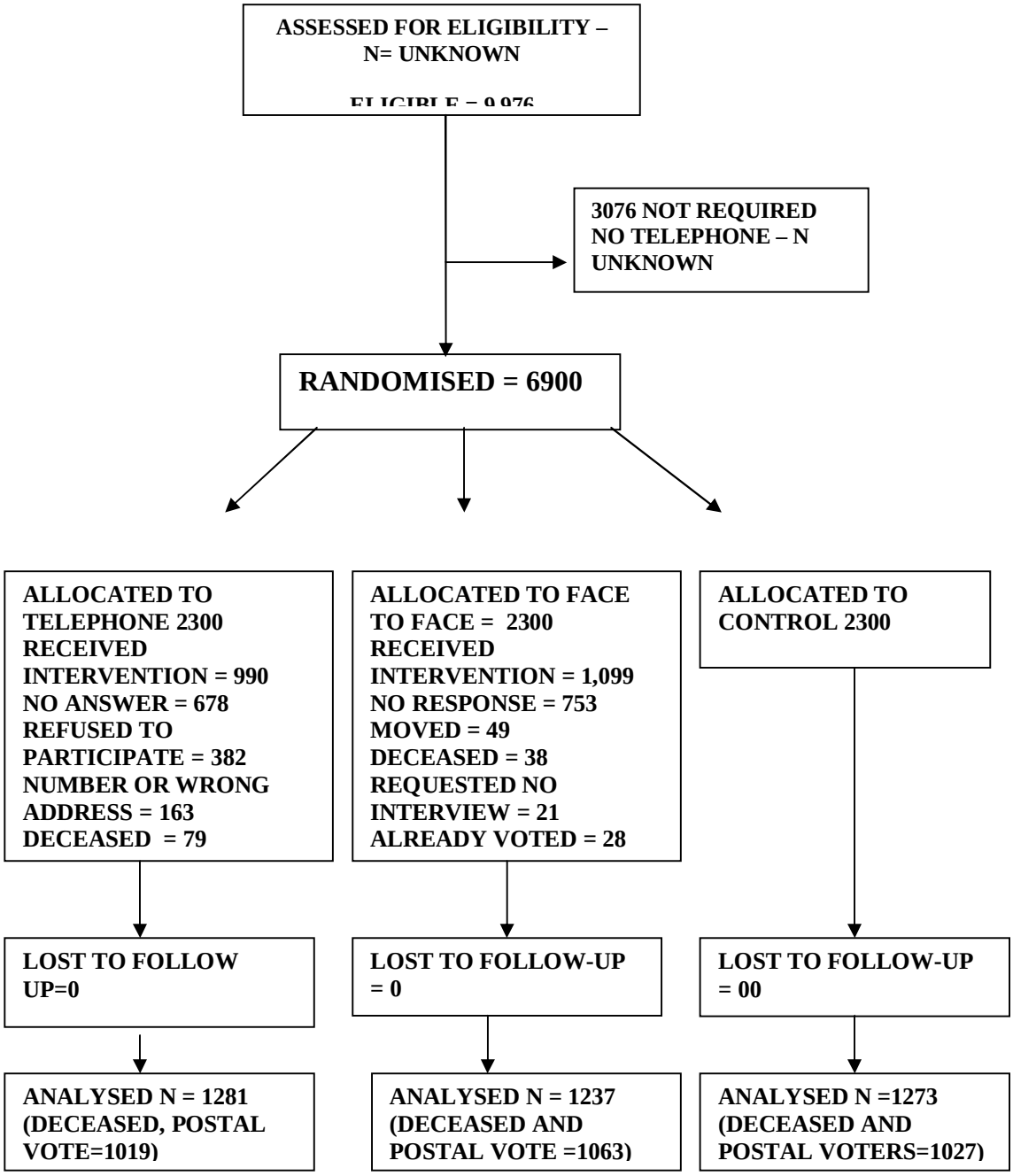
PAPER SECTION and topic	ITEM	Standard CONSORT item Describe	Comment
TITLE & ABSTRACT	1	The participants were randomly allocated using a function in the excel software.	.
INTRODUCTION Background	**2**	To provide evidence on the effectiveness of canvassing in a UK context.	
METHODS Participants	3	Participants had to be on the electoral roll and have a landline telephone number.	The results would not be applicable to people who are 'ex-directory' with 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 non-partisan 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	

PAPER SECTION and topic	ITEM	Standard CONSORT item Describe	Comment
Outcomes	**6**	Main outcome was proportion who voted in the general election. Secondary or process outcomes were the proportion who were successfully contacted.	
Sample size	7	No prior sample size calculation or justification for sample used.	NB with 2,300 in each group the trial would have slightly more than 90% power to show an absolute 5% difference in voting.
Randomization Sequence generation	8	Excel was used to randomise, no detail was given on stratification – probably specified single random samples of 2,300 from overall sample.	
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 and not 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.

PAPER SECTION and topic	ITEM	Standard CONSORT item Describe	Comment
Statistical methods	12	Not clear statistical tests used for the intention to treat 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 estimate total initial sample before exclusions due to lack of telephone land line.	
Recruitment	**14**	Not specified.	
Baseline data	15	Not possible as electoral role gives limited demographical detail of electors.	
Numbers analysed	**16**	In main table of results does not give both numerator and demoninator.	
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 previous, American literature, and shows similar findings.	

PAPER SECTION and topic	ITEM	Standard CONSORT item Describe	Comment
Generalisability	21	May have poor generalisability as it was a single safe constituency in relatively poor area may not apply to wealthier areas.	
Overall evidence	**22**	Draws on past evidence not the answer as results only show a marginal impact.	

FIGURE 1: CONSORT FLOW CHART FOR JOHN AND BRANNAN (2008).



Appendix: Modified Extension of the CONSORT Statement for political science trials adapted from the extension for Nonpharmacologic treatments

PAPER SECTION and topic	ITEM	Standard CONSORT item Describe	Extension for trials of political science trials 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

PAPER SECTION and topic	ITEM	Standard CONSORT item Describe	Extension for trials of political science trials In addition:
	4.B		Details of how the interventions were standardised
	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 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 was 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

PAPER SECTION and topic	ITEM	Standard CONSORT item Describe	Extension for trials of political science trials In addition:
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 interventions1
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

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

PAPER SECTION and topic	ITEM	Standard CONSORT item Describe	Extension for trials of political science trials 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 analysed 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

PAPER SECTION and topic	ITEM	Standard CONSORT item Describe	Extension for trials of political science trials In addition:
Numbers analysed	**16**	Number of participants (denominator) in each group included in each analysis and whether analysis was by "intention-to-treat"; State the results in absolute numbers when feasible (e.g., 10/20, not 50%).	
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% confidence interval)	
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

PAPER SECTION and topic	ITEM	Standard CONSORT item Describe	Extension for trials of political science trials In addition:
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, unequal expertise of intervention providers or centers in each group
Generalisability	21	Generalisability (external validity) of the trial findings	Generalisability (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	