

Using natural experiments to evaluate population health interventions: new Medical Research Council guidance

Peter Craig,¹ Cyrus Cooper,² David Gunnell,³ Sally Haw,⁴ Kenny Lawson,⁵ Sally Macintyre,⁶ David Ogilvie,⁷ Mark Petticrew,⁸ Barney Reeves,⁹ Matt Sutton,¹⁰ Simon Thompson¹¹

► An additional table is published online only. To view this file please visit the journal online (<http://dx.doi.org/10.1136/jech-2011-200375>).

¹MRC Population Health Sciences Research Network and Chief Scientist Office, Scottish Government Health Directorates, Edinburgh, UK

²MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK

³School of Social and Community Medicine, University of Bristol, Bristol, UK

⁴Centre for Public Health and Population Health Research, University of Stirling, Stirling, UK

⁵Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK

⁶MRC CICSO Social and Public Health Sciences Unit, University of Glasgow, Glasgow, UK

⁷MRC Epidemiology Unit and UKCRC Centre for Diet and Activity Research (CEDAR), Cambridge, UK

⁸Department of Social and Environmental Medicine, London School of Hygiene and Tropical Medicine, London, UK

⁹Clinical Trials and Evaluation Unit, University of Bristol, Bristol, UK

¹⁰Health Methodology Research Group, School of Community-based Medicine, University of Manchester, Manchester, UK

¹¹Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Correspondence to

Dr Peter Craig, MRC Population Health Sciences Research Network and Chief Scientist Office, Scottish Government Health Directorates, St Andrews House, Edinburgh EH1 3DG, UK; peter.craig@scotland.gsi.gov.uk

Accepted 26 March 2012

Published Online First
10 May 2012

ABSTRACT

Natural experimental studies are often recommended as a way of understanding the health impact of policies and other large scale interventions. Although they have certain advantages over planned experiments, and may be the only option when it is impossible to manipulate exposure to the intervention, natural experimental studies are more susceptible to bias. This paper introduces new guidance from the Medical Research Council to help researchers and users, funders and publishers of research evidence make the best use of natural experimental approaches to evaluating population health interventions. The guidance emphasises that natural experiments can provide convincing evidence of impact even when effects are small or take time to appear. However, a good understanding is needed of the process determining exposure to the intervention, and careful choice and combination of methods, testing of assumptions and transparent reporting is vital. More could be learnt from natural experiments in future as experience of promising but lesser used methods accumulates.

INTRODUCTION

Natural experimental studies are often recommended as a way of understanding the impact of population-level policies on health outcomes or health inequalities.^{1–4} Within epidemiology there is a long tradition, stretching back to John Snow in the mid nineteenth century,⁵ of using major external shocks such as epidemics, famines or economic crises to study the causes of disease. A difficulty in applying similar methods to the evaluation of population health policies and interventions, such as a 'fat tax' or a legal minimum price per unit of alcohol, is that very often the change in exposure is much less extreme, and its effects may be subtle or take time to emerge. Although they have certain advantages over planned experiments, for example by enabling effects to be studied in whole populations,⁶ and may be the only option when it is impossible to manipulate exposure to the intervention, natural experimental studies are more susceptible to bias and confounding. It is therefore important to be able to distinguish situations in which natural experimental approaches are likely to be informative from those in which some form of fully experimental method such as a randomised controlled trial (RCT) is needed, and from those in which the research questions are genuinely intractable.

The Medical Research Council (MRC) has recently published guidance to help researchers and users, funders and publishers of research evidence make the best use of natural experimental approaches to evaluate population health interventions (<http://www.mrc.ac.uk/naturalexperimentsguidance>). Following the model of the MRC complex interventions guidance,⁷ it was written by a multidisciplinary team with experience of evaluation using a wide range of research designs. The ideas were developed and tested in two specially convened workshops of population health researchers. Drafts were reviewed by workshop delegates and by the MRC's Methodology Research Panel. The guidance is meant to help researchers to plan and design evaluations of public health interventions, journal editors and reviewers to assess the quality of studies that use observational data to evaluate interventions, and policy-makers and others to recognise the strengths and limitations of a natural experimental approach. In this paper we summarise the main messages of the guidance.

WHAT ARE NATURAL EXPERIMENTS?

The term 'natural experiment' lacks an exact definition, and many variants are found in the literature.^{8–10} The common thread in most definitions is that exposure to the event or intervention of interest has not been manipulated by the researcher. Outside an RCT it is rare for variation in exposure to an intervention to be random, so special care is needed in the design, reporting and interpretation of evidence from natural experimental studies, and causal inferences must be drawn with care.

WHY ARE NATURAL EXPERIMENTS IMPORTANT?

Alternatives to RCTs have been advocated by policy-makers and researchers interested in evaluating population-level environmental and non-health sector interventions¹¹ and their impact on health inequalities.⁴ Such interventions may be intrinsically difficult to manipulate experimentally—as in the case of national legislation to improve air quality, or major changes in transport infrastructure¹²—or be implemented in ways that make a planned experiment difficult or impossible, for example with short timescales or extreme variability in implementation.¹³ It may also be unethical to manipulate exposure in order to study effects on

health if an intervention has other known benefits, if it has been shown to be effective in other settings, or if its main purpose is to achieve non-health outcomes.¹⁴ Even if such ethical and practical restrictions are absent, an RCT may still be politically unwelcome.¹⁵

Natural experimental approaches are important because they widen the range of interventions that can usefully be evaluated beyond those that are amenable to planned experimentation. For example, suicide is rare in the general population, occurring at a rate of about 1/10 000 per annum. Even in high risk populations, such as people treated with antidepressants, the annual incidence is only around 1/1000. Clinical trials would have to be enormous to have adequate power to detect even large preventive effects, but natural experiments have been used effectively to assess the impact of measures to restrict access to commonly used means of suicide^{16–18} and inform the content of suicide prevention strategies in the UK and worldwide.¹⁹

These and other studies in which a natural experimental approach has produced clear cut evidence of health impacts are summarised in supplemental table 1. They illustrate the diversity of interventions that have been evaluated as natural experiments, and the wide range of methods that have been applied. Many of the studies have benefited from the availability of high quality, routinely collected data on exposures, potential confounders and outcomes and substantial, rapid changes in exposure across a whole population, which reduces the risk of selective exposure or confounding by secular trends and increases the confidence with which changes in outcomes can be attributed to the interventions. However, it is misleading to assume that whenever a planned experiment is impossible, there is a natural experimental study waiting to happen. Some but not all of the 'multitude of promising initiatives'¹ are likely to yield good natural experimental studies. Care, ingenuity and a watchful eye for good opportunities are needed to realise their potential.

WHEN SHOULD NATURAL EXPERIMENTS BE USED?

The case for adopting a natural experimental approach is strongest when: there is a reasonable expectation that the intervention will have a significant health impact, but scientific uncertainty remains about the size or nature of the effects; an RCT would be impractical or unethical; and the intervention or the principles behind it have the potential for replication, scalability or generalisability.

In practice, natural experiments are highly variable, and researchers face difficult choices about when to adopt a natural experimental approach and how best to exploit the opportunities that do occur. The value of a given natural experiment for research depends on a range of factors including the size of the population affected, the size and timing of likely impacts, the processes generating variation in exposure, and the practicalities of data gathering. Quantitative natural experimental studies should only be attempted when exposed and unexposed populations (or groups subject to varying levels of exposure) can be compared, using samples large enough to detect the expected effects, and when accurate data can be obtained on exposures, outcomes and potential confounders. Resources should only be committed when the economic and scientific rationale for a study can be clearly articulated.

DESIGN, ANALYSIS AND REPORTING OF NATURAL EXPERIMENTS

Planned and natural experiments face some of the same threats to validity, such as loss to follow-up and inaccurate assessment

of exposure and outcomes. The key difference is that RCTs have a very general and (if used properly) effective method of minimising the bias that results from selective exposure to the intervention, that is the tendency for exposure to vary according to characteristics of participants that are also associated with outcomes. In the case of non-randomised studies, there is no such general solution to the pervasive problem of confounding.²⁰ Instead there is a range of partial solutions which can be used in some, often very restricted, circumstances but not others. Understanding the process that produces the variation in exposure (often referred to as the 'assignment process' even when there is no deliberate manipulation of individuals' exposure²¹) is therefore critical to the design of natural experimental studies.⁹

Design

A study protocol should be developed, and ideally published, whatever design is adopted. Good practice in the conduct of observational studies, such as prior specification of hypotheses, clear definitions of target populations, explicit sampling criteria, and valid and reliable measures of exposures and outcomes, should apply equally to natural experimental studies.

All natural experimental studies require a comparison of exposed and unexposed groups (or groups with varying levels of exposure) to identify the effect of the intervention. The examples of suicide prevention,^{16 17} indoor smoking bans^{22–24} and air pollution control^{25 26} show that simple designs can provide convincing evidence if a whole population is abruptly exposed to an intervention, and if the effects are large, rapidly follow exposure and can be measured accurately at population level using routinely available data. This combination of circumstances is rare, and more complex designs are usually required.

Natural experiments can also be used to study more subtle effects, so long as a suitable source of variation in exposure can be found, but the design and analysis of such studies is more challenging. In any case, what is often required is an estimate of effect size, and a large observed effect may incorporate a large element of bias due to selective exposure to the intervention. Whatever the expected effect size, care should be taken to minimise bias in the design and analysis of natural experiments.

Design elements that can strengthen causal inferences from natural experimental studies include the use of multiple pre/post measures to control for secular changes, as in an interrupted time series design²⁷; multiple exposed/unexposed groups that differ according to some variable that may affect exposure and outcome to assess whether selection on that variable is likely to be an important source of bias⁹; accurate measurement of multiple potential confounders and combinations of methods to address different sources of bias. In a study that exemplifies many of the features of a rigorous approach to identifying relatively small effects, Ludwig and Miller²⁸ used variation in access to support for obtaining Headstart funding to model exposure, and compared a variety of outcomes among children who were above or below the age cut-off for access to Headstart services (box 1; supplemental table 1).

Analysis

The defining feature of a natural experiment is that manipulating exposure to the intervention is impossible. There are a few examples where assignment is by a 'real life' lottery, but selection is the rule and a range of methods is available for dealing with the resulting bias.

Where the factors that determine exposure can be measured accurately and comprehensively, matching, regression and

Theory and methods

Box 1 Selection on unobservables

Difference in differences

This method compares change over time in exposed and unexposed groups.³⁶ The differencing procedure controls for unobserved individual differences, and for common trends. It assumes that the unobserved characteristics are fixed, and that the outcomes in each group would change in the same way in the absence of the intervention, so is vulnerable to changes in the composition of the groups and to external influences that differentially affect the exposed and unexposed groups.

Instrumental variables

An instrumental variable is a factor, such as treatment assignment in a well-designed randomised controlled trial, which is associated with outcomes only via its association with exposure to the intervention and is independent of other factors associated with exposure. Instrumental variables have been used to identify the impact of treatment from routine data.^{37–39} In these studies, variables such as distance from a specialised centre have been used to evaluate novel treatments, the assumption being that patients living close to a specialised centre are more likely to receive the novel treatment, but are otherwise similar to other patients.

Regression discontinuity designs

This approach exploits a step change or 'cut-off' in a continuous variable used to assign treatment or otherwise determine exposure to an intervention. The assumption is that units (individuals, areas, etc) just below and just above this threshold will otherwise be similar in terms of characteristics that may influence outcomes, so that an estimate of treatment effect can be obtained by comparing regression slopes on either side of the cut-off. When the Headstart programme to improve the health of disadvantaged children was first implemented in the USA, help with applying for funding was targeted on the 300 poorest counties, and a higher proportion of those counties received funding. Ludwig and Miller²⁸ compared regressions of child mortality on poverty for counties either side of the cut-off, and found lower than expected mortality in those that qualified for assistance.

propensity scores can be used to reduce confounding (box 2). Bias will remain if there are unobserved or imperfectly measured factors that influence both exposure and outcomes. Given the difficulty of measuring accurately all of the characteristics associated with exposure to an intervention, methods such as difference in differences, instrumental variables and regression discontinuity designs that deal with unobserved factors are a potentially valuable advance on those that only deal with observed factors (box 1).

In practice, none of these approaches provides a comprehensive solution to the central problem of selective exposure to the intervention.²⁰ Methods of controlling for observed factors associated with exposure are vulnerable to selection on unobservables. Methods for dealing with selection on unobservables require strong and untestable assumptions²⁹ and their use is restricted by the often very limited availability of variables that can be used to model exposure. They are therefore best used in conjunction with additional tests for the plausibility of any causal inferences.

Box 2 Selection on observables

Matching

This involves finding unexposed individuals (or clusters of individuals) which are similar to those receiving the intervention, and comparing outcomes in the two groups.

Regression adjustment

Measured characteristics that differ between those receiving the intervention and others can be taken into account in multiple regression analyses.

Propensity scores

The likelihood of being exposed to an intervention given a set of covariates can be estimated by logistic regression³⁵ and used to match exposed with unexposed cases, or for covariate adjustment.

Combining methods that address different sources of bias and comparing the results is one such approach and there are several examples in supplemental table 1. In their evaluation of a conditional cash transfer scheme to encourage women to use health facilities to give birth, Lim *et al*³⁰ combined methods for dealing with selection on both observable and non-observable characteristics. Another useful technique is to analyse outcomes that are not expected to change. Dusheiko *et al*³¹ used trends in emergency admissions as a non-equivalent dependent variable to test whether changes in elective admissions could plausibly be attributed to GP fundholding, while Ludwig and Miller²⁸ compared mortality from causes that were likely or unlikely to respond to Headstart services.

Given the difficulty of eliminating bias, single studies are unlikely to be definitive. Replication and careful synthesis of evidence across studies will be needed to support confident inferences about effectiveness. Exact replication of a natural experiment is unlikely, but partial replication is often possible and may be more informative. Consistent findings from studies using varying designs makes it less likely that common biases are present, and consistent findings across settings or populations increase confidence in the generalisability of causal inferences. For example, a number of studies in different countries have shown that legal restrictions on smoking in public places reduce hospital admissions for heart attacks. Although the size of the effect varies widely, as might be expected given variation in smoking rates and the extent of partial restrictions prior to outright bans, the balance of evidence suggests a real effect.²²

Reporting

Transparent reporting of natural experimental studies is vital. Established guidelines such as STROBE³² should be followed, with particular attention to: clearly identifying the approach as a study of a natural experiment; providing a clear description of the intervention and the assignment process; and explicitly stating the methods used to estimate impact. Procedures used to reduce bias should be discussed in a detailed and balanced way. Ideally, qualitative judgements about the risk of bias, and how well it has been dealt with, should be supplemented by a quantitative assessment.^{33 34} If a study has used multiple methods, variation in the estimates should be highlighted. The context within which the intervention was implemented should be described as this may affect interpretation and help users assess the generalisability of the findings. Wherever possible, the results

What is already known on this subject

- Natural experimental approaches widen the range of interventions that can usefully be evaluated, but they are also more prone to bias than randomised controlled trials.
- It is important to understand when and how to use natural experiments and when planned experiments are preferable.

What this study adds

- The UK Medical Research Council has published new guidance on the use of natural experimental approaches to evaluate public health policies and other interventions that affect health.
- Natural experimental approaches work best when the effects of the intervention are large and rapid, and good quality data on exposure and outcomes in a large population are available.
- They can be also used to study more subtle effects, so long as a suitable source of variation in exposure can be found, but the design and analysis of such studies is more demanding.
- Priorities for the future are to build up experience of promising but lesser used methods, and to improve the infrastructure that enables research opportunities presented by natural experiments to be seized.

should be compared with those of other evaluations of similar interventions, paying attention to any associations between effect sizes and variations in evaluation methods and intervention design, content and context.

CONCLUSION

There are important areas of public health policy—such as suicide prevention, air pollution control, public smoking bans and alcohol taxation—where natural experimental studies have already contributed a convincing body of evidence. Such approaches are most readily applied where an intervention is implemented on a large scale, the effects are substantial and good population data on exposure and outcome are available. But they can also be used to detect more subtle effects where there is a suitable source of variation in exposure.

Even so, it would be unwise to assume that a particular policy or intervention could be evaluated as a natural experiment without very detailed consideration of the methodological challenges. Optimism about the use of a natural experimental approach should not be a pretext for discounting the option of conducting a planned experiment, where this would be possible and more robust.

Research effort should focus on addressing important and answerable questions, taking a pragmatic approach based on combinations of research methods and the explicit recognition and careful testing of assumptions. Priorities for the future are to build up experience of promising but lesser used methods, and to improve the infrastructure that enables opportunities presented by natural experiments to be seized, including good routine data from population surveys and administrative sources, good working relationships between researchers and policy makers, and flexible forms of research funding.

Contributors All authors contributed to the planning, drafting and revision of the paper. PC will act as guarantor.

Funding Preparation of this paper was supported by the MRC Population Health Sciences Research Network, and the MRC Methodology Research Panel.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Preparation of the paper did not involve the use of primary data.

REFERENCES

1. Wanless D. *Securing Good Health for the Whole Population*. London: HM Treasury, 2004.
2. Foresight. *Tackling Obesities: Future Choices*. London: Department for Innovation, Universities and Skills, 2007.
3. Department of Health. *Healthy Weight, Healthy Lives: A Research and Surveillance Plan for England*. London: Department of Health, 2008.
4. Kelly MP, Morhan A, Bonnefoy J, et al. *The Social Determinants of Health: Developing An Evidence Base for Political Action. Final Report to World Health Organisation Commission on the Social Determinants of Health*. National Institute for Health and Clinical Excellence. Chile: UK/Universidad de Desarrollo, 2007.
5. Davey Smith G. Commentary: behind the Broad Street pump: aetiology, epidemiology and prevention of cholera in mid 19th century Britain. *Int J Epidemiol* 2002; **31**:920–32.
6. West SG. Alternatives to randomized experiments. *Curr Dir Psych Sci* 2009; **18**:299–304.
7. Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008; **337**:a1655.
8. Academy of Medical Sciences. *Identifying the Environmental Causes of Disease: How Should We Decide What to Believe and When to Take Action*. London: Academy of Medical Sciences, 2007.
9. Meyer BD. Natural and quasi-experiments in economics. *J Bus Econ Stud* 1995; **13**:151–61.
10. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-experimental Designs for Generalised Causal Inference*. Boston, Mass: Houghton Mifflin Company, 2002.
11. West SG, Duan N, Pequegnat W, et al. Alternatives to the randomised controlled trial. *Am J Public Health* 2008; **98**:1359–66.
12. Ogilvie D, Mitchell R, Mutrie N, et al. Evaluating health effects of transport interventions: methodologic case study. *Am J Prev Med* 2006; **31**:118–26.
13. House of Commons Health Committee. *Health Inequalities. Third Report of Session 2008–9. Volume 1 Report, Together with Formal Minutes*. London: The Stationery Office, 2009.
14. Bonell CP, Hargreaves J, Cousens S, et al. Alternatives to randomisation in the evaluation of public health interventions: design challenges and solutions. *J Epidemiol Community Health* 2011; **65**:582–7.
15. Melhuish E, Belsky J, Leyland AH, et al. Effects of fully established Sure Start Local Programmes on 3 year old children and their families living in England: a quasi-experimental observational study. *Lancet* 2008; **372**:1641–7.
16. Gunnell D, Fernando R, Hewagama M, et al. The impact of pesticide regulations on suicide in Sri Lanka. *Int J Epidemiol* 2007; **36**:1235–42.
17. Kretman N. The coal gas story. United Kingdom suicide rates, 1960–71. *Br J Prev Soc Med* 1976; **30**:86–93.
18. Wheeler BW, Gunnell D, Metcalfe C, et al. The population impact on incidence of suicide and non-fatal self harm of regulatory action against the use of selective serotonin reuptake inhibitors in under 18s in the United Kingdom: ecological study. *BMJ* 2008; **336**:542–5.
19. Taylor SJ, Kingdom D, Jenkins R. How are nations trying to prevent suicide? An analysis of national suicide prevention strategies. *Acta Psychiatr Scand* 1997; **95**:457–63.
20. Deeks JJ, Dinnis J, D'Amico R, et al. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003; **7**:1–173.
21. Rubin DR. For objective causal inference, design trumps analysis. *Ann Appl Stat* 2008; **2**:808–40.
22. Mackay D, Pell J, Irfan M, et al. Meta-analysis of the influence of smoke-free legislation on acute coronary events. *Heart* 2010; **96**:1525–30.
23. Pell J, Haw SJ, Cobbe S, et al. Smokefree legislation and hospitalisations for acute coronary syndrome. *N Engl J Med* 2008; **359**:482–91.
24. Sims M, Maxwell R, Bauld L, et al. Short term impact of smoke-free legislation in England: retrospective analysis of hospital admissions for myocardial infarction. *BMJ* 2010; **340**:c2161.
25. Clancy L, Goodman P, Sinclair H, et al. Effect of air pollution control on death rates in Dublin, Ireland: an intervention study. *Lancet* 2002; **360**:1210–14.
26. Hedley AJ, Wong CM, Thach TQ, et al. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: an intervention study. *Lancet* 2002; **360**:1646–52.
27. Campostrini S, Holtzman D, McQueen D, et al. Evaluating the effectiveness of health promotion policy: changes in the law on drinking and driving in California. *Health Promot Int* 2006; **21**:130–5.

Theory and methods

28. **Ludwig J**, Miller DL. Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *Q J Econ* 2007;122:159–208.
29. **Hernan MA**, Robins JM. Instruments for causal inference. An epidemiologist's dream? *Epidemiology* 2006;17:360–72.
30. **Lim SS**, Dandona L, Hoisington JA, et al. India's Janani Suraksha Yonana: a conditional cash transfer scheme to increase births in health facilities: an impact evaluation. *Lancet* 2010;375:2009–23.
31. **Dusheiko M**, Gravelle H, Jacobs R, et al. *The Effect of Budgets on Doctor Behaviour: Evidence from a Natural Experiment*. Discussion Papers in Economics. York: University of York, 2003.
32. **von Elm E**, Altman D, Egger M, et al: STROBE Initiative. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;335:806–8.
33. **Fox MP**. Creating a demand for bias analysis in epidemiological research. *J Epidemiol Community Health* 2009;63:91.
34. **Turner RM**, Spiegelhalter DJ, Smith GC, et al. Bias modelling in evidence synthesis. *J R Stat Soc Ser A Stat Soc* 2009;172:21–47.
35. **D'Agostino RB**. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998;17:2265–81.
36. **Angrist J**, Pischke JS. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press, 2009.
37. **Brookhart MA**, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010;19:537–54.
38. **McClellan M**, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA* 1994;272:859–66.
39. **Stukel TA**, Fisher ES, Wennberg DE, et al. Analysis of observational studies in the presence of treatment selection bias. Effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. *JAMA* 2007;297:278–85.
40. **Herttua K**, Makela P, Martikainen P. Changes in alcohol-related mortality and its socioeconomic differences after a large reduction in alcohol prices: a natural experiment based on register data. *Am J Epidemiol* 2008;168:1110–18.
41. **Herttua K**, Makela P, Martikainen P. An evaluation of the impact of a large reduction in alcohol prices on alcohol-related and all-cause mortality: time series analysis of a population-based natural experiment. *Int J Epidemiol* 2011;40:441–54.