Making better use of natural experimental evaluation in population health

Peter Craig, professor[1], Mhairi Campbell, research associate[1] Adrian Bauman, professor emeritus[2] Manuela Deidda, research fellow[3] Ruth Dundas, professor[1] Niamh Fitzgerald, professor[4] Judith Green, professor [5] Srinivasa Vittal Katikireddi, professor [1] Jim Lewsey, professor [3] David Ogilvie, professor[6] Frank de Vocht, professor [78] Martin White professor [6]

[1]MRC/CSO Social and Public Health Sciences Unit, School of Health and Wellbeing, University of Glasgow, Glasgow, UK

[2]School of Public Health, Charles Perkins Centre, University of Sydney, Sydney, Australia

[3]Health Economics and Health Technology Assessment, School of Health and Wellbeing, University of Glasgow, Glasgow, UK

[4]Institute for Social Marketing and Health, University of Stirling, Stirling, UK

[5]Wellcome Centre for Cultures and Environments of Health, University of Exeter, Exeter, UK

[6]MRC Epidemiology Unit, University of Cambridge, Cambridge, UK

[7]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

[8]NIHR Applied Research Collaboration West

Correspondence to: M Campbell Mhairi.Campbell@glasgow.ac.uk

Rather than arguing about the suitability of natural experimental methods to inform decisions we need to focus on refining their scope and design, say ***Peter Craig and colleagues***

Key messages

- Natural experimental evaluations can provide useful information to guide decision making about interventions
- Most discussion has focussed on what quantitative methods are suitable for natural experimental evaluations

Key definitions and concepts remain contested and there is a lack of consensus about the circumstances in which natural experimental evaluations can provide trustworthy and useful evidence for decision-making

- Guidance should help identify the circumstances that make for good natural experimental evaluation, and a range of applicable methods

Natural experiments have long been used as opportunities to evaluate the health impacts of policies, programmes, and other interventions. Defined in the UK Medical Research Council's guidance as events outside the control of researchers that divide populations into exposed and unexposed groups, natural experiments have greatly contributed to the evidence base for tobacco and air pollution control, suicide prevention, and other important areas of public health policy.[1]

Although randomised controlled trials are often viewed as the best source of evidence because they have less risk of bias, reliance on them as the only source of credible evidence has begun to shift for several reasons. Firstly, policy makers are increasingly looking for evidence about "what works" to tackle pervasive and complex problems, including the social determinants of health,[2 3] and these are hard to examine in randomised trials. In Scotland, for example, legislation to introduce a minimum retail price per unit of alcohol included a sunset clause, which means that the measure will lapse after six years unless evidence is produced that it works. This has resulted in multiple evaluations, including natural experimental studies using geographical or historical comparator groups.[4] Similarly, the US National Institutes of Health has called for greater use of natural experimental methods to understand how to prevent obesity,[5] and a consortium of European academies for their greater use to understand policies and interventions to reduce health inequalities.[3]

Secondly, a wider range of analytical methods developed within other disciplines, mostly by economists or other social or political scientists, are being increasingly applied to good effect. A good example is the use of synthetic control methods to evaluate the effect on mortality of the introduction of a pay-for-performance scheme for financing hospital care.[6] There is also a greater availability of large administrative and other "big" data sources that link information on exposure to public policies with health and other outcomes.

Although natural experimental evaluations have an established foothold in population health research, particularly to support policy making, more work is needed to identify the best opportunities for natural experimental studies and to support their design, conduct, and synthesis to realise their full potential.

Diversifying the sources of evidence

The idea that there is a hierarchy of study designs, ranked according to susceptibility to bias, remains influential.[7 8] A common shorthand for this view is that randomised controlled trials are the gold standard for evaluation, and that observational study designs are irredeemably weaker in all circumstances. An alternative view is that, while unbiased

estimates of effectiveness are an important goal of evaluations, they are not the only goal, and may be unachievable in some circumstances.[9] If research seeks to produce evidence that is useful for policy and other decision making, a wider range of study designs is needed, including those that work in situations when a planned experiment would not be feasible or ethical.

During the covid-19 pandemic, for example, randomised trials provided crucial evidence about the efficacy of vaccines in reducing risk of infection and severe disease, as well as the efficacy of treatments.[10 11] But observational studies of the effects of interventions in practice have also made important contributions[12 13]: evidence on longer term effectiveness of vaccines, has been established through cohort studies conducted in the context of large scale vaccination campaigns when it was unethical to withhold vaccines that had already been shown to be efficacious[12] and evidence on the impact of physical distancing interventions was generated employing interrupted time series analyses using routine data from 149 countries.[13] Furthermore, evidence on the effectiveness of surveillance was obtained from ingenious natural experimental evaluations that exploited flaws in the implementation of the test and trace programme in England.[14 15] The adverse effects of the UK Treasury's "eat out to help out" scheme on infection rates were likewise identified by treating the scheme as a natural experiment.[16]

Despite growing acceptance of the value of natural experimental evalutions,[17-20] most discussion of their design focuses on quantitative methods.[17 18 21] Other aspects of study design, conduct, and interpretation, such as how to identify a good opportunity for a natural experimental evaluation, how open science principles such as registering a study should be applied, or how to place effect estimates into a broader framework of whether, how, and in what circumstances the intervention achieves its effects, have been largely neglected.

Moving from classifications to opportunities

One key conceptual issue is how broadly or narrowly natural experiments should be defined.[22] A related question is whether it is useful to distinguish sharply on methodological grounds between natural experimental evaluations and other observational studies that attempt to identify causal relationships using change or variation in exposure that is not associated with a specific event or process, such as the implementation of a new policy.

The UK MRC's broad definition of a natural experiment contrasts with attempts to narrow the definition to include only studies that use one of a prescribed range of analytical methods,[23] or that satisfy some other criterion such as "as if randomisation" (a real world

process leading to variation in exposure that approximates random allocation in a trial,[24] such as the use of lotteries to allocate military conscription[25] or school places.[26] However, broad study design labels are an inadequate proxy for study quality, which depends on the extent to which assumptions are tested, threats to validity evaluated, and robustness checks performed.[27] Lists of approved analytical methods can rapidly become dated as new methods are developed and existing ones refined. For example, synthetic control methods, which use a weighted composite of control areas rather than a geographical control area, have been widely applied to evaluate public health and healthcare interventions such as state level tobacco[28] and firearms control[29] policies in the US, but rarely feature in such lists.[30]

Additionally, even though "as if randomisation" provides a strong basis for causal inference from a natural experiment, other than in clear-cut cases such as lotteries it is difficult to define precisely when the criterion is satisfied: few population health interventions are or could be implemented in this way. Rather than trying to sharply differentiate natural experimental evaluations from other observational studies on methodological grounds, it may be more useful to think about the sets of circumstances that are likely to generate useful opportunities for robust research using natural experimental evaluations.[14]

Recognising opportunities for natural experimental studies

Natural experimental evaluations are most commonly used in situations where there is a clear division in presence, level, time, place, or type of exposure between two or more otherwise similar subpopulations—for example, when a policy is implemented in one state within a federal jurisdiction but not in neighbouring states. Several other situations recur in the literature and provide useful pointers for the design of future studies, including policies with eligibility criteria that clearly define exposure, phased implementation of policy, the use of randomisation to determine entitlements or obligations, and flaws or shortcomings in policy delivery (table 1).

Table 1 Policy-related opportunities for natural experimental evaluations

Type of opportunity	Examples
New policy with a clearly defined inception date or administrative jurisdiction	Gun control laws are implemented at state level within the US, so the effect of introducing or withdrawing a law in one state can be evaluated using other states as controls[29]
Policy that applies to some members of a population but not others	Social security policies often define eligibility for payment in terms of age or income, allowing comparisons between individuals with ages or incomes just above or just below the eligibility cut-off[31]

Phased implementation of a policy across a population	Universal Credit, a new system of social security benefits and tax credits in the UK, was rolled out area by area, allowing comparisons of recipients of benefits under the old system and the new system using data from surveys running throughout the rollout period[32 33]
Policies using randomisation as an assignment mechanism, generating otherwise similar exposed and unexposed individuals or groups	Election laws in two Indian states required a minimum number of council seats be reserved for women in a randomly selected proportion of villages, allowing comparisons of decision making in villages with larger or smaller numbers of women involved[34]
Flaws or shortcomings in policy implementation	Database errors led to people in some areas being randomly excluded from England's test and trace programme for covid-19 until the error was noticed. This enabled trends in infection rates to be compared in areas with and without tracing in operation[14]

Whether such situations generate opportunities for good evaluations depends as much on availability and quality of data as on the nature of events or processes themselves.[22] Many natural experimental evaluations are conducted retrospectively so good quality, routinely collected data from administrative systems, population surveys, or other sources is critical. Similarly, it is important to be able to accurately characterise the nature and timing and, where relevant, the intensity and implementation of the intervention being evaluated to correctly identify individuals or groups who were or were not exposed, or had varying levels of exposure. This often relies on access to good quality documentary evidence, as well as access to key informants who can remember and reliably describe the intervention, including how and when it was implemented.

Study registration policies

Another question that has been relatively neglected with respect to natural experimental evaluations is the registration of study protocols. For prospective studies such as randomised controlled trials, registration, especially if enforced by funders and stipulated by journal editors, is a powerful safeguard against some forms of manipulation, such as selective publication of favourable findings. For retrospective studies, where researchers may be familiar with a dataset before the study begins, transparency about how such prior knowledge has affected design choices is vital.

The protocols for natural experimental evaluations may have to be amended to accommodate changes as the evaluation progresses, such as developments in theoretical understanding of the nature of the intervention or a fuller appreciation of the characteristics of the data. For example, missing data may require modification of the analysis plan to use a different set of covariates or an alternative analytic method. Often protocols are published in a journal with no facility for updating. Natural experimental evaluation protocols may benefit

from the flexibility now accepted for systematic review protocols, which can have amendments recorded.[35 36]

Improved evidence base

Evidence from natural experimental evaluations can provide insight beyond estimates of effect size and contribute to understanding the importance of mechanisms or context of interventions within systems.[37 38] Greater awareness and use of a range of methods to estimate the effects of interventions not under researchers' control—and to understand how, where, and for whom those effects are realised—is essential for developing a robust and useful evidence base for policy. Indeed, natural experimental evaluations have already proved their value across a wide and disparate range of health and non-health policy areas, providing otherwise unobtainable evidence about the effects on population health of clean air legislation, suicide prevention, tobacco and gun control, trade agreements, non-pharmaceutical pandemic control measures, and many other kinds of interventions.

We believe they can contribute much further to these and other areas if the focus moves beyond justifying their use to optimising their execution. Making the most of evidence that can be obtained from natural experiments requires incorporating economic evaluations and modelling[39 3] as well as qualitative methods that could provide vital information about possible causal mechanisms.[40 41] Further guidance on how to identify opportunities for natural experimental evaluations, on how to design, conduct, report, and synthesise the evidence from such studies, and on what kinds of research infrastructure and governance processes are needed will help to realise this potential.

Contributors and sources: PC led the writing of the article and is the guarantor. MC, AB, MD, RD, NF, JG, JL, DO, SVK, FdV, and MW contributed to the conceptualisation and commented on successive drafts and approved the final version.

Acknowledgments: The project is funded by the National Institute for Health Research and Medical Research Council (MC_PC_21009). PC, MC, RD and SVK are supported by the Medical Research Council (MC_UU_00022/2) and the Scottish Government Chief Scientist Office (SPHSU17). SVK is supported by an NHS Research Scotland Senior Clinical Fellowship (SCAF/15/02). RD is supported by UK Prevention Research Partnership (MR/S037608/1). NF is supported by grants from NIHR, CSO and UKRI unrelated to this paper. JG is supported by Wellcome Trust (Centre Grant 203109/Z/16/Z). DO is supported by the Medical Research Council (Unit programme MC_UU_00006/7). MW is supported by the Medical Research Council (Unit programme MC_UU_00006/7), and grants from UKRI and NIHR unrelated to this paper.

Competing interests: We have read and understood BMJ policy on declaration of interests, and declare support from the NIHR and MRC for work related to this article.

Provenance and peer review: Not commissioned; externally peer reviewed.

References

1 Craig P, Cooper C, Gunnell D, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. J Epidemiol Community Health 2012;66:1182-6. PubMed doi:10.1136/jech-2011-200375

2 Snell A, Reeves A, Rieger M, et al. WHO Regional Office for Europe's natural experiment studies project: an introduction to the series. Eur J Public Health 2018;28(suppl_2):1-3. PubMed doi:10.1093/eurpub/cky195

3 European Federation of Academies of Sciences and Humanities (ALLEA), Federation of European Academies of Medicine (FEAM). Health inequalities research. New methods, better insights? Royal Netherlands Academy of Arts and Sciences, 2021. https://allea.org/wp-content/uploads/2021/11/Health_Inequalities.pdf

4 Robinson M, Mackay D, Giles L, Lewsey J, Richardson E, Beeston C. Evaluating the impact of minimum unit pricing (MUP) on off-trade alcohol sales in Scotland: an interrupted time-series study. Addiction 2021;116:2697-707. PubMed doi:10.1111/add.15478

5 National Institutes of Health Office of Disease Prevention. A report from the federal partners meeting of the national institutes of health pathways to prevention workshop: methods for evaluating natural experiments in obesity. 2018. https://prevention.nih.gov/sites/default/files/2019-01/ObesityMethodsP2PFederalPartnersMeetingReport.pdf

6 Kreif N, Grieve R, Hangartner D, Turner AJ, Nikolova S, Sutton M. Examination of the Synthetic Control Method for Evaluating Health Policies with Multiple Treated Units. Health Economics. 2016;25(12):1514-28.

7 Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. Soc Sci Med 2018;210:2-21. PubMed doi:10.1016/j.socscimed.2017.12.005

8 Rothman KJ. Six persistent research misconceptions. J Gen Intern Med 2014;29:1060-4. PubMed doi:10.1007/s11606-013-2755-z

9 Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021;374:n2061. PubMed doi:10.1136/bmj.n2061

10 Ramasamy MN, Minassian AM, Ewer KJ, et al; Oxford COVID Vaccine Trial Group. Safety and immunogenicity of ChAdOx1 nCoV-19 vaccine administered in a prime-boost

regimen in young and old adults (COV002): a single-blind, randomised, controlled, phase 2/3 trial. Lancet 2021;396:1979-93. PubMed doi:10.1016/S0140-6736(20)32466-1
11 Horby P, Lim WS, Emberson JR, et al; RECOVERY Collaborative Group. Dexamethasone in hospitalized patients with covid-19. N Engl J Med 2021;384:693-704. PubMed doi:10.1056/NEJMoa2021436
12 Agrawal U, Katikireddi SV, McCowan C, et al. Covid-19 hospital admissions and deaths after BNT162b2 and ChAdOx1 nCoV-19 vaccinations in 2·57 million people in Scotland (EAVE II): a prospective cohort study. Lancet Respir Med 2021;9:1439-49. PubMed doi:10.1016/S2213-2600(21)00380-5
13 Islam N, Sharp SJ, Chowell G, et al. Physical distancing interventions and incidence of coronavirus disease 2019: natural experiment in 149 countries. BMJ 2020;370:m2743. PubMed doi:10.1136/bmj.m2743
14 Fetzer T. Measuring the epidemiological impact of a false negative: evidence from a natural experiment. University of Warwick, Department of Economics, 2021.
15 Fetzer T, Graeber T. Measuring the scientific effectiveness of contact tracing: evidence from a natural experiment. Proc Natl Acad Sci U S A 2021;118:e2100814118. PubMed doi:10.1073/pnas.2100814118
16 Fetzer T. Subsidising the spread of covid-19: evidence from the UK's eat-out-to-help-out scheme. Econ J (Lond) 2021;132:1200-17.
17 Basu S, Meghani A, Siddiqi A. Evaluating the health impact of large-scale public policy changes: classical and novel approaches. Annu Rev Public Health 2017;38:351-70. PubMed doi:10.1146/annurev-publhealth-031816-044208
18 Craig P, Katikireddi SV, Leyland A, Popham F. Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. Annu Rev Public Health 2017;38:39-56. PubMed doi:10.1146/annurev-publhealth-031816-044327
19 Matthay EC, Glymour MM. Causal inference challenges and new directions for epidemiologic research on the health effects of social policies. Curr Epidemiol Rep 2022;9:22-37. doi:10.1007/s40471-022-00288-7.
20 Matthay EC, Hagan E, Gottlieb LM, et al. Alternative causal inference methods in population health research: evaluating tradeoffs and triangulating evidence. SSM Popul Health 2019;10:100526. PubMed doi:10.1016/j.ssmph.2019.100526
21 Bärnighausen T, Tugwell P, Røttingen J-A, et al. Quasi-experimental study designs series-paper 4: uses and value. J Clin Epidemiol 2017;89:21-9. PubMed doi:10.1016/j.jclinepi.2017.03.012
22 de Vocht F, Katikireddi SV, McQuire C, Tilling K, Hickman M, Craig P. Conceptualising natural and quasi experiments in public health. BMC Med Res Methodol 2021;21:32. PubMed doi:10.1186/s12874-021-01224-x
23 Tugwell P, Knottnerus JA, McGowan J, Tricco A. Big-5 quasi-experimental designs. J Clin Epidemiol 2017;89:1-3. PubMed doi:10.1016/j.jclinepi.2017.09.010
24 Dunning T. Natural experiments in the social sciences: a design-based approach. Cambridge University Press, 2012 doi:10.1017/CBO9781139084444.
25 Angrist JD. Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records. Am Econ Rev. 1990;80:313–36.
26 Angrist J, Bettinger E, Bloom E, King E, Kremer M. Vouchers for private schooling in Colombia: evidence from a randomized natural experiment. Am Econ Rev. 2002;92:1535–58
27 Reeves BC, Wells GA, Waddington H. Quasi-experimental study designs series-paper 5: a checklist for classifying studies evaluating the effects on health interventions-a taxonomy without labels. J Clin Epidemiol 2017;89:30-42. PubMed doi:10.1016/j.jclinepi.2017.02.016

28 Abadie A, Diamond A, Hainmueller J. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. J Am Stat Assoc. 2010;105:493–505
29 Humphreys DK, Gasparrini A, Wiebe DJ. Evaluating the impact of Florida's "stand your ground" self-defense law on homicide and suicide by firearm: an interrupted time series study. JAMA Intern Med 2017;177:44-50. PubMed doi:10.1001/jamainternmed.2016.6811
30 Bouttell J, Craig P, Lewsey J, Robinson M, Popham F. Synthetic control methodology as a tool for evaluating population-level health interventions. J Epidemiol Community Health 2018;72:673-8. PubMed doi:10.1136/jech-2017-210106
31 Katikireddi SV, Molaodi OR, Gibson M, Dundas R, Craig P. Effects of restrictions to Income Support on health of lone mothers in the UK: a natural experiment study. Lancet Public Health 2018;3:e333-40. PubMed doi:10.1016/S2468-2667(18)30109-9
32 Wickham S, Bentley L, Rose T, Whitehead M, Taylor-Robinson D, Barr B. Effects on mental health of a UK welfare reform, universal credit: a longitudinal controlled study. Lancet Public Health 2020;5:e157-64. PubMed doi:10.1016/S2468-2667(20)30026-8
33 Brewer M, Dang T, Tominey E. Universal Credit: Welfare Reform and Mental Health. IZA Institute of Labor Economics, 2022.
34 Chattopadhyay R, Duflo E. Women as policy makers: evidence from a randomized policy experiment in India. Econometrica 2004;72:1409-43 doi:10.1111/j.1468-0262.2004.00539.x.
35 Centre for Reviews and Dissemination. Guidance notes for registering a systematic review protocol with PROSPERO. University of York, 2016.
36 Lasserson TJ, Thomas J, Higgins JP. Starting a review. In: Cochrane handbook for systematic reviews of interventions. 2021. www.training.cochrane.org/handbook.
37 Ogilvie D, Adams J, Bauman A, et al. Using natural experimental studies to guide public health action: turning the evidence-based medicine paradigm on its head. J Epidemiol Community Health 2020;74:203-8. PubMed doi:10.1136/jech-2019-213085
38 White M, Adams J. Different scientific approaches are needed to generate stronger evidence for population health improvement. PLoS Med 2018;15:e1002639. PubMed doi:10.1371/journal.pmed.1002639
39 Deidda M, Geue C, Kreif N, Dundas R, McIntosh E. A framework for conducting economic evaluations alongside natural experiments. Soc Sci Med 2019;220:353-61. PubMed doi:10.1016/j.socscimed.2018.11.032
40 Hanckel B, Petticrew M, Thomas J, Green J. The use of qualitative comparative analysis (QCA) to address causality in complex systems: a systematic review of research on public health interventions. BMC Public Health 2021;21:877. PubMed doi:10.1186/s12889-021-10926-2
41 Green J, Roberts H, Petticrew M, et al. Integrating quasi-experimental and inductive designs in evaluation: a case study of the impact of free bus travel on public health. Evaluation 2015;21:391-406 doi:10.1177/1356389015605205.