



OPEN ACCESS

# Using natural experimental studies to guide public health action: turning the evidence-based medicine paradigm on its head

David Ogilvie ,<sup>1</sup> Jean Adams,<sup>1</sup> Adrian Bauman,<sup>2</sup> Edward W. Gregg,<sup>3</sup> Jenna Panter,<sup>1</sup> Karen R. Siegel,<sup>4</sup> Nicholas J. Wareham,<sup>1</sup> Martin White<sup>1</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jech-2019-213085>).

<sup>1</sup>MRC Epidemiology Unit and Centre for Diet and Activity Research (CEDAR), University of Cambridge, Cambridge, UK

<sup>2</sup>Charles Perkins Centre and Prevention Research Collaboration, University of Sydney, Sydney, New South Wales, Australia

<sup>3</sup>School of Public Health, Imperial College, London, UK

<sup>4</sup>National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

## Correspondence to

Dr David Ogilvie, MRC Epidemiology Unit and Centre for Diet and Activity Research (CEDAR), University of Cambridge, Cambridge, UK; [david.ogilvie@mrc-epid.cam.ac.uk](mailto:david.ogilvie@mrc-epid.cam.ac.uk)

Received 15 August 2019

Revised 30 October 2019

Accepted 31 October 2019

## ABSTRACT

Despite smaller effect sizes, interventions delivered at population level to prevent non-communicable diseases generally have greater reach, impact and equity than those delivered to high-risk groups. Nevertheless, how to shift population behaviour patterns in this way remains one of the greatest uncertainties for research and policy. Evidence about behaviour change interventions that are easier to evaluate tends to overshadow that for population-wide and system-wide approaches that generate and sustain healthier behaviours. Population health interventions are often implemented as natural experiments, which makes their evaluation more complex and unpredictable than a typical randomised controlled trial (RCT). We discuss the growing importance of evaluating natural experiments and their distinctive contribution to the evidence for public health policy. We contrast the established evidence-based practice pathway, in which RCTs generate 'definitive' evidence for particular interventions, with a practice-based evidence pathway in which evaluation can help adjust the compass bearing of existing policy. We propose that intervention studies should focus on reducing critical uncertainties, that non-randomised study designs should be embraced rather than tolerated and that a more nuanced approach to appraising the utility of diverse types of evidence is required. The complex evidence needed to guide public health action is not necessarily the same as that which is needed to provide an unbiased effect size estimate. The practice-based evidence pathway is neither inferior nor merely the best available when all else fails. It is often the only way to generate meaningful evidence to address critical questions about investing in population health interventions.

## INTRODUCTION

Governments around the world are committed to tackling the growing burden of non-communicable diseases.<sup>1</sup> Unhealthy patterns of behaviours such as smoking, diet, alcohol consumption and physical activity contribute substantially to disease risk and life expectancy, particularly in middle-income and high-income countries.<sup>2</sup> The populations of lower-income countries are increasingly at risk of similar behaviour patterns as their economies develop. For example, insufficient physical activity is already estimated to account for nearly 10% of global premature mortality.<sup>3</sup> The energy imbalance and metabolic health of people in lower-income countries are liable to worsen with the increasing

availability of high-energy ultraprocessed foods and mechanisation of labour and transport that characterise an aspirational 'Western' lifestyle.

The aetiological associations between behaviours and many chronic disease outcomes are sufficiently well established to justify efforts to ameliorate behavioural risk factors. However, how population behaviour patterns might most effectively be shifted remains one of the greatest uncertainties for public health research and policy. To date, effort has largely been directed at developing and evaluating interventions to change the behaviours of individuals at higher risk, sometimes successfully. However, it is doubtful that merely scaling up these approaches to reach more and more people would be affordable, effective or equitable as a global disease prevention strategy.<sup>2 4–6</sup> A more authentic and sustainable population-based strategy would complement the current focus on effective primary and secondary prevention—targeting individuals at higher risk—with more primordial prevention (table 1) that addresses the environments and policies that shape the circumstances in which we live. This vision has deep roots in the histories of public health and medicine.<sup>7</sup> However, we remain largely ignorant about how to achieve it. This ignorance reflects the challenges in evaluating primordial prevention strategies of this kind, interpreting the findings and translating them into action. These challenges are reflected in the conclusions of recent systematic reviews, as exemplified in table 2.

In this paper, we discuss the growing importance of evaluating natural experiments in primordial prevention and their distinctive contribution to generating evidence for public health policy. We identify some of the obstacles to this type of research and suggest greater effort and investment in this area to ensure that research more effectively supports public health action.

## WHAT DO WE NEED TO KNOW?

We should reflect on the extent to which our research is aligned with the societal processes it is intended to inform.<sup>8</sup> To that end, we must distinguish between 'behaviour change' as a population goal on the one hand, that is, the *outcome* we ultimately wish to achieve; and on the other, 'behaviour change' as an intervention strategy or moral imperative, that is, a way of framing both the problem (people are making poor behavioural choices) and the solution (they need to make different choices). An approach



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY. Published by BMJ.

**To cite:** Ogilvie D, Adams J, Bauman A, et al. *J Epidemiol Community Health* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jech-2019-213085

**Table 1** Glossary of selected terms

| Term                          | Meaning in this paper  |
|-------------------------------|--|
| Decision-theoretical approach | 'A decision-theory approach utilizes relevant knowledge, theory and data both from (sic) observational and experimental studies to evaluate the likely efficacy of an intervention. If from this process it can be demonstrated that an intervention is sufficiently unlikely to cause net harm, then we can move to estimate cost-effectiveness. That is, we assess if the benefit relative to its cost is sufficient for the intervention to be recommended for application to population groups under consideration. This contrasts with the hypothesis-testing approach in which decisions about the efficacy of an intervention are made solely by using the findings of scientific studies that use statistical testing to evaluate their efficacy. The hypothesis-testing approach is central to evidence-based medicine but in practice groups charged with reaching decisions about health interventions for populations also use additional evidence alongside scientific, methodological and philosophical judgements.' <sup>27</sup> |
| Natural experiment            | 'The term(...)lacks an exact definition, and many variants are found in the literature. The common thread in most definitions is that exposure to the event or intervention of interest has not been manipulated by the researcher.' <sup>18</sup><br>'Natural experiments are, by definition, events that occur outside the control of the researcher. They are not 'conducted' or 'designed'; on the contrary, they are discovered.' <sup>22</sup>   |
| Primordial prevention         | 'This term is advocated by some authors to describe elimination of risk factors(...)in contrast to primary prevention by reducing risks of exposure.' <sup>(S1)</sup>  |
| Quasi-experiment              | 'A situation in which the investigator lacks full control over the allocation and/or timing of intervention but nonetheless conducts the study as if it were an experiment, allocating subjects to groups. Inability to allocate subjects randomly is a common situation that may be best described as a quasi-experiment.' <sup>(S1)</sup>  |

Citation S1 in this table refers to the online supplementary reference list.

too narrowly focused on people and their behaviour or lifestyle as the problem, and thereby on interventions that often seek 'to persuade the poor to change their behaviour',<sup>9</sup> is not compatible with a social ecological understanding of the causes of ill health that are not amenable to individual control.<sup>2 10</sup>

The evidence available to guide policy has long been subject to an 'evaluative bias' in favour of behavioural interventions targeting people at higher risk because such interventions are generally easier to evaluate and, in particular, easier to randomise.<sup>11</sup> This type of evidence tends to overshadow that for strategies that act on whole populations by targeting critical leverage points in the *systems* that generate and sustain less healthy behaviour patterns.<sup>2 12</sup> This implies a need to direct greater policy and research attention to where the underlying problems are located: not merely among individuals at higher risk, nor even among groups of more deprived individuals, but in the more fundamental causes of 'dis-ease' (sic)<sup>13</sup> in communities—for example, in the unhealthy environments created as a consequence of the structural conditions of the planning, transport and welfare systems and the housing and labour markets.<sup>10</sup>

**WHY DO WE NOT KNOW?**

The lack of evidence for effective primordial prevention strategies may be traced to one of three types of obstacle.

The first is a set of *political* obstacles encountered by researchers who are willing but unable to produce the evidence. Researchers seeking to evaluate environmental or policy interventions—such as improving access to green space or taxing

particular foods—depend on governments or other agencies to implement evaluable strategies.<sup>6</sup> Because these interventions often entail greater political cost or risk than those focused on individual choice, they tend to be introduced less often. When such policies do find favour, demonstration projects and similar initiatives are often introduced quickly, without time to establish rigorous evaluation studies.<sup>14</sup> Even if an intervention is both promising and evaluable in principle,<sup>15</sup> an agile evaluative response may depend on more rapid and flexible sources of funding than have traditionally been available.

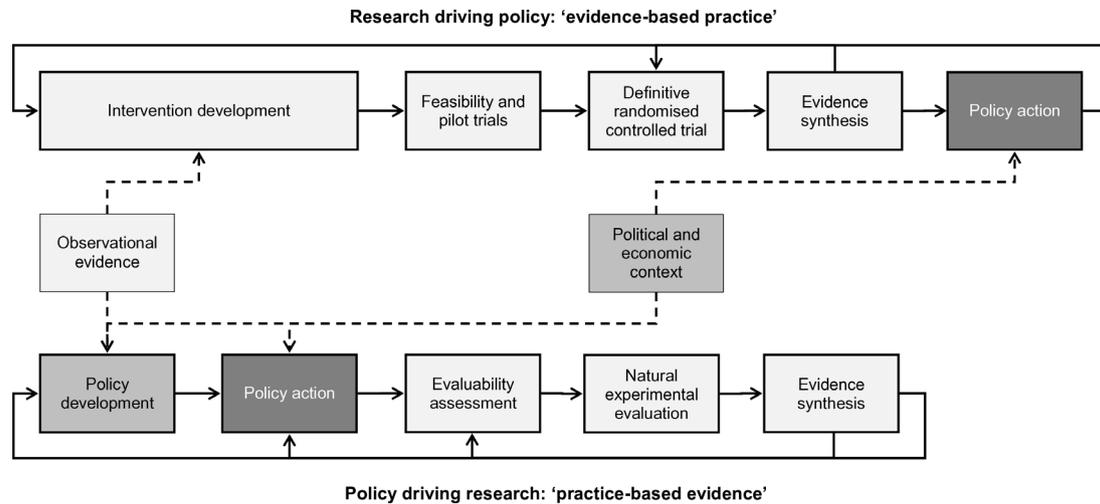
The second is a set of *cultural* obstacles in research, manifested by a research community that is able in principle to produce the evidence, but rendered somewhat unwilling by circumstances. Nearly two decades ago, it was pointed out that only a small fraction of UK public health research expenditure was directed towards 'solutions'.<sup>16</sup> Today, observational epidemiology and the development and evaluation of targeted behaviour change interventions remain easier and more secure routes to 'doing something', achieving funding, producing publications and career progression.<sup>12 17</sup> A research community that, quite understandably, 'follows the money' in this way may therefore be distorting the agenda in research (and, consequently, in policy).

The third is a set of *practical* obstacles. Primordial preventive strategies are generally implemented as 'natural' or 'quasi-' experiments rather than 'true' experiments (table 1).<sup>18</sup> Evaluating these strategies thereby makes for a more complex and unpredictable undertaking than, for example, a typical clinical trial—which is not without enormous potential challenges of its

**Table 2** Key findings of recent examples of systematic reviews

| Topic   | Key points from main results and authors' conclusions (emphases added)   |
|---|--|
| Interventions to reduce ambient particulate matter air pollution and their effect on health                             | 'The evidence base, comprising non-randomized studies only, was of <b>low or very low certainty</b> (...)Given the heterogeneity across interventions, outcomes, and methods, it was <b>difficult to derive overall conclusions</b> (...)The evidence base highlights the <b>challenges related to establishing a causal relationship</b> between specific air pollution interventions and outcomes(...)Researchers should strive to sufficiently account for confounding, assess the impact of methodological decisions(...)and improve the reporting of methods, and other aspects of the study, most importantly the description of the intervention and the context in which it is implemented.' <sup>(S2)</sup> |
| Fortification of staple foods with vitamin A for vitamin A deficiency   | ' <b>We are uncertain</b> whether fortifying staple foods with vitamin A alone makes little or no difference for serum retinol concentration(...)It is <b>uncertain</b> whether vitamin A alone reduces the risk of subclinical vitamin A deficiency(...)The certainty of the evidence was mainly affected by <b>risk of bias, imprecision and inconsistency</b> .' <sup>(S3)</sup>  |
| Nutritional interventions for preventing stunting in children living in urban slums in low- and middle-income countries | 'Overall, the <b>evidence was complex</b> to report, with a wide range of outcome measures reported(...)The <b>certainty of evidence was very low to moderate</b> (...)All the nutritional interventions reviewed had the <b>potential</b> to decrease stunting(...)however, there was <b>no evidence of an effect</b> of the interventions included in this review(...)More evidence is needed of the effects of multi-sectorial (sic) interventions(...)as well as the effects of <b>'up-stream' practices and policies</b> ' <sup>(S4)</sup>  |
| Environmental interventions to reduce the consumption of sugar-sweetened beverages and their effects on health          | ' <b>We judged most studies to be at high or unclear risk of bias</b> in at least one domain, and most studies used non-randomised designs(...)Implementation should be accompanied by <b>high-quality evaluations using appropriate study designs, with a particular focus on the long-term effects</b> of approaches suitable for large-scale implementation.' <sup>(S5)</sup>   |
| Fortification of wheat and maize flour with folic acid for population health outcomes                                   | ' <b>Most studies had unclear risk of bias</b> (...)Limitations of this review were the small number of studies and participants, <b>limitations in study design, and low-certainty of evidence due to how included studies were designed and reported</b> .' <sup>(S6)</sup>  |

Source: The five most recently published systematic reviews listed on the website of the Cochrane Public Health Group (<http://ph.cochrane.org/cph-reviews-and-topics>, accessed 4 July 2019). Citations in this table refer to the online supplementary reference list.



**Figure 1** Two complementary modes of evidence generation.

own. This calls for a particularly flexible and nuanced approach to natural experimental study design and analysis, along with sufficient capability and capacity to deliver this.<sup>6</sup> No wonder, then, that it seems more common to see papers calling for this type of research than to see papers reporting it.

## TWO COMPLEMENTARY MODES OF EVIDENCE GENERATION

Much has been written about the bench-to-bedside translational medicine pathway linking basic science with clinical practice. That concept has also strongly influenced thinking about evidence to support public health action, for example, in discourse that refers to institutionalising ‘proven’, ‘evidence-based’ interventions.<sup>19</sup> It envisages a largely unidirectional pipeline in which researchers—informed by observational studies—develop interventions, subject them to feasibility and pilot testing, and then evaluate them in definitive randomised controlled trials (RCTs). These are often conducted in settings and groups of people unlike those in which a public health intervention might ultimately be applied. When a systematic review of multiple trials concludes that an intervention is effective, that intervention is regarded as ‘proven’ and may be recommended by a body such as the National Institute for Health and Care Excellence (NICE: [www.nice.org.uk](http://www.nice.org.uk)) for more widespread implementation, subject to broader contextual considerations such as affordability and political acceptability. This is shown in the upper pathway of [figure 1](#). In this pathway, the purpose of evaluation can be seen as indicating whether a traffic light holding back an ‘unproven’ intervention should be turned from red to green.

However, this implied linear, rational way in which new knowledge is converted into ‘evidence-based practice’ has limited empirical support or practical utility for ‘upstream’ public health interventions.<sup>19</sup> Even in comparatively well-resourced healthcare systems, major preventive initiatives such as cervical screening and routine health checks for over-40s have been introduced without evidence of effectiveness from RCTs to support them.<sup>4,20</sup> This applies even more outside healthcare, where actions influencing complex systems of wider determinants of health such as food supply, income and urban planning are being taken all the time. These actions occur for a variety of reasons that may or may not be ostensibly concerned with health, and with or without evidence to support them or meaningful evaluation to learn from them.<sup>2,6</sup> For example, when a new neighbourhood is built to accommodate unmet need for housing, planners make decisions about the mix of land uses, amenities provided (such

as schools and parks) and street network layout. Each of these decisions constitutes an intervention that may influence physical activity and the risk of chronic disease among the people who live there.<sup>21</sup> It is, however, unlikely to be realistic for governments to take no action until ‘sufficient’ evidence of all such effects have been cumulated and synthesised from multiple intervention studies.<sup>22,23</sup>

Public health research is concerned with ‘generating discoveries and new knowledge within the public health field itself’.<sup>8</sup> This implies an opportunity and need to complement the evidence-based practice pathway, described above, with innovative solutions—generated in and for the real world by policy-makers and practitioners—that can also be rigorously evaluated to produce ‘practice-based evidence’ (as shown in the lower pathway of [figure 1](#)).<sup>5</sup> In this latter pathway, the purpose of evaluation can be seen as more akin to adjusting the compass bearing followed by existing policy rather than enabling a binary decision to proceed or not with the widespread implementation of a particular intervention. Whereas the former pathway depends on multiple instances of evidence of *effectiveness* to justify action, the latter depends more on multiple instances of *action* from which to develop at least preliminary evidence of effectiveness. This may in turn support taking further evaluable action and the consequent, cumulative reduction of uncertainty about its effects.

The application of each step of the practice-based evidence pathway, and how it differs from the converse pathway, can be illustrated with a worked example based on the published protocol for the ongoing evaluation of the UK soft drinks industry levy,<sup>24</sup> a fiscal measure intended to reduce consumption ([table 3](#)). This particular intervention had not been implemented before, so the case for action was not based on established evidence of effectiveness as such. Rather, it rested on a plausible<sup>25</sup> case for effectiveness based on a combination of observational epidemiology, simulation modelling and a limited set of evaluations of related interventions in other parts of the world. Without a political decision to ‘do something’ based on such ‘non-randomised’ evidence, it would be impossible ever to generate stronger evidence about effectiveness. Thoughtful assessment of the evaluability of the intervention revealed the complexity of its theorised mechanisms and potential outcomes, along with the unfeasibility of imposing an RCT design on this particular fiscal policy measure in this particular context. This dictated a natural experimental evaluation using a combination

**Table 3** Natural experimental evaluation of the UK treasury soft drinks industry levy

| Step in pathway                           | Précis of protocol (published in 2017) <sup>24</sup>   |
|---|--|
| Observational evidence                    | Sugar-sweetened beverage (SSB) consumption is independently associated with multiple chronic disease outcomes. SSBs currently represent the single biggest source of dietary sugar for young people in the UK. Economic theory and data strongly suggest that price is an important determinant of SSB purchases.  |
| Policy development                        | To reduce population consumption of SSBs, a range of interventions had been proposed, including fiscal measures. Globally a number of SSB taxes had been introduced, although few had been evaluated when this study was initiated. Modelling studies suggested important potential health gains, but no comprehensive evaluations had measured impacts on reformulation or consumption.   |
| Policy action                             | In 2016, the Chancellor of the Exchequer announced a tiered soft drinks industry levy (SDIL) on industries importing or selling SSBs in the UK with the explicit intention of reducing consumption of sugar from SSBs. At the time of announcement, the UK SDIL was different from other SSB taxes: it was an industry levy (paid by manufacturers) rather than an excise tax (paid by consumers).   |
| Evaluability assessment                   | The implementation of a fiscal policy is an intervention that is highly context dependent, resulting in reactions by many stakeholders including government, civil society, industry, health sector and consumers, and the potential to affect a range of diet and health outcomes. The SDIL was unique in its construction including a tiered levy directed at industry, and its 2-year lead time from date of announcement to implementation. Randomised controlled trials are recognised as the strongest method for determining causal effects. However, in the current context where the SDIL was introduced to the whole country at once, randomisation to intervention and control groups was not feasible.   |
| Design of natural experimental evaluation | This evaluation seeks to improve our understanding of how such interventions evolve over time within complex food systems to influence products and purchasing, consumption and health outcomes. The evaluation will thus take a 'systems' perspective, aiming to evaluate a range of outcomes, associated processes and their dynamic interrelationships. Interrupted time series (ITS) methods offer one of the strongest quasi-experimental research designs. Using ITS designs, consideration of a range of outcomes (eg, SSB consumption declining as consumption of lower sugar alternatives increase) and mechanistic processes (eg, the relationship between price and purchases) can be explored such that a 'pattern' of impacts is appraised to provide the strongest possible basis on which to draw causal inference. |
| Evidence synthesis                        | Findings will be integrated and synthesised to develop a coherent overarching interpretation (and) test and refine the underlying intervention theory for the SDIL. Findings generated using different methods (qualitative, quantitative) could be triangulated to explore the extent to which they provide a consistent interpretation and conclusions about the impacts of the SDIL using pattern matching and causal process observation, thus strengthening causal inference.   |

of *most appropriate* methods (eg, interrupted time series analysis) to systematically rule out alternative explanations for observed effects and demonstrate credible causal pathways leading to those effects.<sup>25</sup> In contrast to the reliance on successful randomisation for causal attribution in an RCT, this study relies on integrating the findings of multiple quantitative and qualitative components for deriving robust inferences. These will contribute to subsequent evidence synthesis, as much in terms of validating (potentially generalisable) overall intervention theory as in terms of producing (context-specific) effect size estimates for meta-analysis.<sup>6,26</sup> Such findings can be used to adjust existing policies, or inform future actions around the world, in this area to optimise their health outcomes.

**TOWARDS A GOOD ENOUGH EVIDENCE BASE FOR PUBLIC HEALTH ACTION**

How, then, might effort and investment in developing the published evidence base more effectively support the kinds of policy intervention required for primordial prevention? Our analysis suggests three main implications; examples of potential actions arising from these are given in [table 4](#).

**Intervention studies should focus on reducing critical uncertainties**

Studies of interventions to change upstream determinants of disease risk, such as population dietary or commuting patterns,

are sometimes criticised because they have not followed participants to ultimate, 'hard' physiological or clinical endpoints. However, there is no reason to expect that all parts of a putative causal chain should be directly proved within a single study. It has been argued that to inform action, public health intervention research should be guided as much by a decision-theoretical approach ([table 1](#)) as by the narrower, but more familiar, statistical hypothesis-testing approach.<sup>27</sup> This implies that evaluation should focus on reducing the most critical uncertainties<sup>28</sup> about what should be done—that is, the ways in which various intervention strategies influence population *behaviour* patterns—just as we judge smoking cessation services not on their direct impact on heart disease or lung cancer, but on whether they help people quit smoking.<sup>29</sup> A case for (or against) action can be gradually cumulated using the iterative exchange of data and theory between empirical observational and intervention studies in a variety of contexts, simulation modelling of more distal or long-term health impacts and other sources of evidence.<sup>30</sup>

**Non-randomised study designs should be embraced rather than tolerated**

Although natural experimental study designs have important theoretical underpinnings in common with the RCT, their worth does not reside solely in the extent to which they emulate an RCT design. Dunning proposes three criteria for assessing the utility of natural experimental studies.<sup>31</sup> The first is that the allocation

**Table 4** Examples of potential actions to help develop the evidence base

| Arena                | Potential action   |
|----------------------|--|
| Research training    | The teaching of more established research methods in observational epidemiology, randomised controlled trials and meta-analysis in postgraduate training could be complemented with other mixed-method and natural experimental approaches more often taught and used in the social sciences |
| Research methodology | Efforts to develop consensus and guidance on the appraisal of natural experimental studies could be expanded, complementing the current emphasis on internal validity with greater consideration of external validity, transferability and utility for informing action                      |
| Research funding     | Research funding bodies and their peer reviewers could assess natural experimental studies more closely on their own merits rather than using templates based too closely on the expectations of a typical randomised controlled trial   |
| Academic publishing  | Journals could adopt editorial policies committed to selecting manuscripts based more on the applicability of a given study design to a given research question than on prior assumptions about a hierarchy of study design  |
| Policymaking         | Policymakers could allow more time (and assign funding, if appropriate) to enable adequate theorisation, robust study design and baseline data collection to be undertaken before new policies and other interventions are implemented   |
| Knowledge exchange   | Policy and research communities could establish horizon-scanning or intelligence-sharing networks to bring implementers and potential evaluators into dialogue as early as possible in the process of establishing new interventions   |

of an intervention can be treated ‘as if it were random, although not within a planned RCT. Although randomisation eliminates important sources of potential confounding, an expectation that intervention studies should entail a comparable allocation process (such as a lottery) would further entrench existing evaluative biases because many interventions relevant to public health are never likely to fulfil this criterion.<sup>22</sup> This may be because randomisation is impractical (eg, new transport infrastructure is built in particular places for particular reasons) or politically unpalatable (if, eg, it is seen as withholding a service from certain areas or groups).<sup>12</sup> Furthermore, intervention studies that ‘fail’ this criterion may pass with flying colours on Dunning’s other two criteria for utility. One of these relates to the *relevance* of the intervention to current, real-world policy questions. A key advantage of natural experimental studies is that they ‘do not interfere in the natural data generation process’,<sup>32</sup> and thereby largely avoid the problems of ‘artificial and less directly informative’ inferences from effects observed in experimental studies in more controlled settings.<sup>33</sup> The other criterion relates to the *plausibility* of the causal inference.<sup>25</sup> Here again, a natural experimental study may be ‘more likely to generate causal evidence that applies to intervention implementation in real life’,<sup>34</sup> particularly if it elicits evidence of *how* an intervention achieves its effects.<sup>27</sup>

Of course, this may appear to sit uneasily within a research funding system based on a biomedical paradigm that privileges the RCT above all other methods for establishing effectiveness.<sup>35</sup> But randomisation does not necessarily hold the key to unlocking questions about public health action.<sup>25</sup> Nor does the proliferation of epidemiological studies that link environmental exposures with health behaviours in a statistically robust way but are incapable of testing whether altering the former influences the latter.<sup>21 36</sup> If a given method or study design is chosen for its alignment with the applied research question and executed in a rigorous and transparent way, it is likely to contribute important evidence even though (and perhaps *because*) it falls into the implicitly disparaging category of ‘non-randomised’ studies.<sup>35</sup>

### A more thoughtful approach to appraising the utility of evidence

This is not to deny that many non-randomised studies do have major limitations and are reported in ways that lack rigour or transparency. For example, systematic reviews of studies linking changes in the built environment with changes in diet, physical activity and adiposity have noted multiple potential sources of bias and that ‘studies with weaker designs were more likely to report associations in the positive direction’.<sup>23 37</sup> In addition to all the issues that complicate the practice and interpretation of trials, in a natural experimental study close attention needs to be paid to understanding exactly what exposure to an intervention consists of; how an intervention comes to be assigned to some people, groups or areas and not others; finding a valid basis for estimating the counterfactual, such as by using a meaningful control group or a graded measure of intervention exposure; selecting and interpreting the adjustment for appropriate covariates to minimise the risk of confounding; and interpreting complex patterns within the outcomes, which may include divergent and potentially inequitable responses between subgroups, dose-response relationships and comparisons with multiple controls.<sup>18 38 39</sup>

We have well-established, and continually developing, catechisms for assessing the internal validity of intervention studies, and groups of studies, in health research.<sup>40</sup> However, we lack clear consensus on the relative importance or interpretation of different

aspects of internal validity in natural experimental studies, and therefore on how to make constructive use of an evidence base that fits poorly into existing appraisal systems.<sup>6</sup> For example, current tools for assessing risk of bias appear predicated on a preference for studies that resemble an RCT as closely as possible.<sup>22 23</sup> They tend to downplay or ignore the importance of ‘greater qualitative appraisal (and) theoretical and statistical knowledge’,<sup>32</sup> and of what different quantitative and qualitative components of single or multiple studies might contribute in combination to a growing body of overall, more generalisable causal inference.<sup>31 33</sup> In particular, we lack consensus on ‘how good is good enough’—which partly depends, of course, on the answer to the question ‘good enough for what?’ The complex evidence needed to guide public health action is not necessarily the same as that which is needed to provide an unbiased estimate of an effect size.

### CONCLUSION

We are more likely to halt the rise in the global prevalence of non-communicable diseases by taking and evaluating new, more ambitious or radical actions to address the underlying causes than by merely applying existing preventive approaches—even if these are effective—with greater intensity. Even apparently simple questions about effectiveness in this arena cannot be answered without action—although based on the best available evidence at the time—that necessarily precedes evaluation. The practice-based evidence pathway can be regarded as an essential, and currently under-resourced and undervalued, complement to the more established evidence-based practice pathway. It is neither inferior nor merely the best available when all else fails. On the contrary, it is often the only way to generate meaningful evidence to address critical questions about investing in population health interventions.

#### What is already known on this subject

- ▶ There are well-established associations between behaviour and chronic disease, which justify government efforts to reduce behavioural risk factors. However, the question of how population behaviour patterns might most effectively be shifted remains one of the greatest uncertainties for research and policy. This reflects the substantial challenges of evaluating population preventive strategies, interpreting the findings and translating them into action. Greater effort and investment in this area may help ensure that research more effectively supports public health action.

#### What this study adds

- ▶ We discuss the growing importance of evaluating natural experiments and their distinctive contribution to the evidence for public health policy. We contrast the established evidence-based practice pathway, in which randomised controlled trials generate ‘definitive’ evidence for particular interventions, with a practice-based evidence pathway in which evaluation can help adjust the compass bearing of existing policy. We propose that intervention studies should focus on reducing critical uncertainties, that non-randomised study designs should be embraced rather than tolerated and that a more nuanced approach to appraising the utility of diverse types of evidence is required.

The two pathways for generating evidence described in this paper do not represent mutually exclusive approaches. Some policy and practice innovations could and should be evaluated in RCTs, and many more would benefit from more planned evaluation using a wider range of study designs. Nevertheless, the public health research community and those who fund and publish their work have key roles to play in supporting the development and credibility of researchers in this field, and the more thoughtful conduct, appraisal and synthesis of natural experimental studies, to populate critical missing pieces of the evidence base to support more effective public health action.

**Contributors** DO conceived of the original idea and drafted the initial manuscript. JA, AB, EWG, JP, KRS, NJW and MW provided critical feedback and contributed to the final version of the manuscript. DO is the guarantor.

**Funding** DO, JP and NJW are supported by the Medical Research Council (Unit Programme numbers MC\_UU\_12015/6 and MC\_UU\_12015/1). The paper was initially developed in the course of a visiting appointment as Thought Leader in Residence at the School of Public Health at the University of Sydney, for which the intellectual environment and financial support provided by the Prevention Research Collaboration is gratefully acknowledged. It was further developed under the auspices of the Centre for Diet and Activity Research (CEDAR), a UKCRC Public Health Research Centre of Excellence at the University of Cambridge, for which funding from the British Heart Foundation, Economic and Social Research Council, Medical Research Council, National Institute for Health Research and the Wellcome Trust, under the auspices of the United Kingdom Clinical Research Collaboration, is gratefully acknowledged; and through the authors' collaboration in organising a workshop on the evaluation of natural experiments of social and environmental interventions with potential impacts on population risk of diabetes and cardiometabolic disease in Atlanta on 7–8 March 2017, at which much of the content was presented and for which funding from the Centers for Disease Control and Prevention is gratefully acknowledged.

**Disclaimer** The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or other funders mentioned.

**Competing interests** None declared.

**Patient consent for publication** Not required.

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

**Data availability statement** There are no data in this work.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

#### ORCID iD

David Ogilvie <http://orcid.org/0000-0002-0270-4672>

#### REFERENCES

- General Assembly of the United Nations. *Political Declaration of the high-level meeting of the general assembly on the prevention and control of non-communicable diseases*. New York: General Assembly of the United Nations, 2011.
- Marteau TM, White M, Rutter H, et al. Increasing healthy life expectancy equitably in England by 5 years by 2035: could it be achieved? *The Lancet* 2019;393:2571–3.
- Lee I-M, Shiroma EJ, Lobelo F, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The Lancet* 2012;380:219–29.
- Capewell S, McCartney M, Holland W. NHS health checks – a naked emperor? *J Public Health* 2015;37:187–92.
- Reis RS, Salvo D, Ogilvie D, et al. Scaling up physical activity interventions worldwide: stepping up to larger and smarter approaches to get people moving. *The Lancet* 2016;388:1337–48.
- White M, Adams J. Different scientific approaches are needed to generate stronger evidence for population health improvement. *PLoS Med* 2018;15:e1002639.
- Mackenbach JP. Politics is nothing but medicine at a larger scale: reflections on public health's biggest idea. *J Epidemiol Commun Health* 2009;63:181–4.
- Viehbeck SM, Petticrew M, Cummins S. Old myths, new myths: challenging myths in public health. *Am J Public Health* 2015;105:665–9.
- Green J. What kind of research does public health need? *Crit Public Health* 2014;24:249–52. [10.1080/09581596.2014.917813](https://doi.org/10.1080/09581596.2014.917813)
- Panter-Brick C, Eggerman M, Tomlinson M. How might global health master deadly SINS and strive for greater virtues? *Glob Health Action* 2014;7:23411.
- Ogilvie D, Egan M, Hamilton V, et al. Systematic reviews of health effects of social interventions: 2. best available evidence: how low should you go? *J Epidemiol Community Health* 2005;59:886–92.
- Rutter H, Savona N, Glonti K, et al. The need for a complex systems model of evidence for public health. *The Lancet* 2017;390:2602–4.
- Krieger N. Commentary: ways of asking and ways of living: reflections on the 50th anniversary of Morris' ever-useful uses of epidemiology. *Int J Epidemiol* 2007;36:1173–80.
- House of Commons Health Committee. *Health inequalities. third report of session 2008–09*. London: Stationery Office, 2009.
- Ogilvie D, Cummins S, Petticrew M, et al. Assessing the evaluability of complex public health interventions: five questions for researchers, funders, and policymakers. *Milbank Q* 2011;89:206–25.
- Millward L, Kelly M, Nutbeam D. *Public health intervention research: the evidence*. London: Health Development Agency, 2001.
- Academy of Medical Sciences. *Improving the health of the public by 2040: optimising the research environment for a healthier, fairer future*. London: Academy of Medical Sciences, 2016.
- 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.
- Ogilvie D, Craig P, Griffin S, et al. A translational framework for public health research. *BMC Public Health* 2009;9:116.
- Cochrane AL, Holland WW. Validation of screening procedures. *Br Med Bull* 1971;27:3–8.
- Sallis JF, Cerin E, Conway TL, et al. Physical activity in relation to urban environments in 14 cities worldwide: a cross-sectional study. *The Lancet* 2016;387:2207–17.
- Humphreys DK, Panter J, Ogilvie D. Questioning the application of risk of bias tools in appraising evidence from natural experimental studies: critical reflections on Benton et al., IJBNPA 2016. *Int J Behav Nutr Phys Act* 2017;14.
- Benton JS, Anderson J, Hunter RF, et al. The effect of changing the built environment on physical activity: a quantitative review of the risk of bias in natural experiments. *Int J Behav Nutr Phys Act* 2016;13.
- White M. *Evaluation of the health impacts of the UK Treasury Soft Drinks Industry Levy (SDIL) [protocol]*. National Institute for Health Research: Southampton, 2017.
- Victoria CG, Habicht J-P, Bryce J. Evidence-Based public health: moving beyond randomized trials. *Am J Public Health* 2004;94:400–5.
- Panter J, Guell C, Prins R, et al. Physical activity and the environment: conceptual review and framework for intervention research. *Int J Behav Nutr Phys Act* 2017;14.
- Threlfall AG, Meah S, Fischer AJ, et al. The appraisal of public health interventions: the use of theory. *J Public Health* 2015;37:166–71.
- Ling T. Evaluating complex and unfolding interventions in real time. *Evaluation* 2012;18:79–91.
- Lancaster T, Stead LF, Cochrane Tobacco Addiction Group. Individual behavioural counselling for smoking cessation. *Cochrane Database Syst Rev* 2017;101.
- Diez Roux AV. Complex systems thinking and current impasses in health disparities research. *Am J Public Health* 2011;101:1627–34.
- Dunning T. *Natural experiments in the social sciences: a design-based approach*. Cambridge: Cambridge University Press, 2012.
- Waddington H, Aloe AM, Becker BJ, et al. Quasi-Experimental study designs series—paper 6: risk of bias assessment. *J Clin Epidemiol* 2017;89:43–52.
- Geng EH, Peiris D, Kruk ME. Implementation science: relevance in the real world without sacrificing rigor. *PLoS Med* 2017;14:e1002288.
- Bärnighausen T, Tugwell P, Rottingen J-A, et al. Quasi-Experimental study designs series—paper 4: uses and value. *J Clin Epidemiol* 2017;89:21–9.
- Petticrew M, Roberts H, Evidence RH. Evidence, hierarchies, and typologies: horses for courses. *J Epidemiol Community Health* 2003;57:527–9.
- Bauman AE, Reis RS, Sallis JF, et al. Correlates of physical activity: why are some people physically active and others not? *The Lancet* 2012;380:258–71.
- Mayne SL, Auchincloss AH, Michael YL. Impact of policy and built environment changes on obesity-related outcomes: a systematic review of naturally occurring experiments. *Obes Rev* 2015;16:362–75.
- Trochim WMK. Outcome pattern matching and program theory. *Eval Program Plann* 1989;12:355–66.
- Humphreys DK, Panter J, Sahlqvist S, et al. Changing the environment to improve population health: a framework for considering exposure in natural experimental studies. *J Epidemiol Community Health* 2016;70:941–6.
- Higgins J, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0: Cochrane Collaboration*, 2011.