



COMMENTARY

Quasi-experiments are a valuable source of evidence about effects of interventions, programs and policies: commentary from the Campbell Collaboration Study Design and Bias Assessment Working Group

Hugh Sharma Waddington^{a,*}, David B. Wilson^b, Terri Pigott^c, Gavin Stewart^d, Ariel M. Aloe^e, Peter Tugwell^f, Vivian Welch^g

^aLondon School of Hygiene and Tropical Medicine, Department of Disease Control, Environmental Health Group, UK

^bCriminology, Law and Society Department, George Mason University, Fairfax, VA, USA

^cSchool of Public Health and the College of Education & Human Development, Georgia State University, Atlanta, GA, USA

^dModelling, Evidence & Policy Group, Newcastle University, UK

^eCollege of Education, University of Iowa, Iowa City, IA, USA

^fFaculty of Medicine, Department of Epidemiology & Community Medicine, Bruyère Research Institute, University of Ottawa, Ottawa, Ontario, Canada

^gThe Campbell Collaboration, Oslo, Norway

Accepted 2 November 2022; Published online xxxx

We have read with interest the guidelines paper by Saldanha and coauthors [1] for the Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Center (EPC) Program. The paper aims to articulate the circumstances in which systematic reviews of health interventions should incorporate nonrandomized studies of interventions. These are studies that aim to demonstrate, and quantify, the causal effect of a defined treatment (an intervention, program, or policy) on a defined outcome, where the allocation to the treatment condition uses some method other than randomization. Examples of nonrandom allocation mechanisms include assignment based on a threshold on a pretest score, as in the case of the regression discontinuity design (RDD). Alternatively, participants may be selected by planners and practitioners using some other method (for example, by administrative jurisdiction), and/or it may simply be that study participants self-select into treatment, and the selection rules modelled in data analysis.

The authors use the terminology “nonrandomized studies of interventions (NRSI)”, popularized in Sterne et al. [2], which incorporates nonrandomized studies of effects designed prospectively (that is, trials) or retrospectively (that is, observational studies). Following accepted terminology in the social sciences and related fields, we refer to these types of studies as quasi-experimental designs (QEDs) [3].

Saldanha et al. have updated previous guidance from AHRQ, which had argued that NRSI should be considered admissible in systematic reviews for harm outcomes, where randomized controlled trials (RCTs) are likely to be insufficient. The authors have switched emphasis to focus on the potential contribution of NRSI as opposed to the rationale of only including them when the evidence base is otherwise lacking. They state that “instead of recommending that NRSIs be included only if RCTs are insufficient to address the key question, or that NRSIs always be included, the current guidance considers NRSIs as potentially important.” (p.15)

The paper is of high topic relevance. For example, a recent report from the US National Academy of Sciences [4] calls for educational research to focus on heterogeneity of treatment effects, or, in other words, to design studies to examine why, how and for whom interventions work, or do not. The report argues that the current focus on the internal validity of education intervention studies has limited the ability of policymakers to apply these results to contexts different from those represented in the RCT. Evidence from a wider range of contexts and participants provides a deeper understanding of both the average treatment effect and how it might vary.

As is commonly voiced in policy research, the methods applied should follow from the review question and not the other way round. Drawing on Saldanha and colleagues stated reasons, we list below why we believe QEDs provide a crucial source of evidence in reviews of intervention effects. We note that there may be material differences in studies of effects in health care and in social policy, such as in isolating the effect of a medical treatment from placebo. There may also be areas and questions where large, well conducted RCTs are more plentiful, and

Declaration: The opinions are our own and do not necessarily represent the views of the Campbell Collaboration.

Conflict of interest: The authors are not aware of any interests, financial or otherwise, which might affect our commentary.

* Corresponding author. London International Development Centre, 20 Bloomsbury Square, London WC1A 2NS, UK. Tel.: +44 20 7958 8251.

E-mail address: hugh.waddington@lshtm.ac.uk (H. Sharma Waddington).

What is new?**Key findings**

- Quasi-experimental designs (QEDs), also called nonrandomized studies of intervention effects, can provide evidence that is both internally and externally valid for decision making.

What this adds to what is known

- There are stronger theoretical and empirical reasons to incorporate QEDs in systematic reviews of intervention effects than are commonly acknowledged.

What is the implication, what should change now

- Systematic reviews of intervention effects should usually incorporate appropriately critically-appraised evidence from QEDs.

correspondingly where research on strong QEDs has been traditionally more limited.

1. The external validity argument: QEDs provide evidence on the effectiveness of interventions in a wide variety of fields (e.g., health, social policy, economic development, and the environment). One reason why they are invaluable is that they tend to provide evidence of intervention effects conducted under circumstances of usual treatment practice, rather than in circumstances where allocation has been modified by researchers or treatments implemented with a higher degree of fidelity than is usually seen in practice. That is, QEDs provide evidence on treatment effectiveness under “real world” conditions, rather than treatment efficacy as a proof-of-concept. Evidence on intervention effectiveness is important for decision makers, who, in most instances, are interested not just in the population average effect, but in understanding the variability in findings (for whom, in what circumstances, why). A key advantage of incorporating QEDs in meta-analysis is that they can facilitate exploration of this variability. For example, network meta-analysis will frequently include open loops when restricted to RCTs; adding in appropriately critically appraised QEDs may provide the missing connectivity evidence to close these loops. As is the case with RCTs, the critical appraisal of QEDs is especially important to avoid incorporating biased estimates, which sometimes operate to produce overly optimistic effects.

2. The internal validity argument (theoretical): Saldanha et al. note threats to the internal validity of non-randomized studies, including confounding, selection and misclassification biases, but there are also threats to validity in prospective studies like RCTs, which are not applicable to observational studies that are designed retrospectively. An example is motivational effects: when study participants are aware they are being observed as part of trial, this might affect their motivation, either to improve their performance or to demoralize them [5]. Thus, inclusion of different study designs could allow a triangulation approach to assessment of internal validity, where different designs contribute different information to different elements of both internal and external validity. Bradford-Hill [6] referred to this as “consistency.”
3. The internal validity argument (empirical): the evidence cited by Saldanha et al. to support inclusion of nonrandomized studies is from meta-epidemiology. This evidence is susceptible to confounding—by population, intervention, comparator, outcome and setting—because the comparisons are indirect. In other words, the comparisons made of RCTs and NRSI/QEDs in these meta-analyses are from external replications that do not necessarily use the same populations at the same time periods or the same intervention and counterfactual conditions. Evidence on empirical validation from internal replication studies (also called within-study comparisons) is thought more credible as it directly compares the QED and RCT findings in the same populations, time-periods and settings. Systematic reviews and meta-analyses of these studies (e.g., [7,8]) support the internal validity of NRSI/QEDs, especially for strong designs like RDD. These studies can also provide empirical bias correction factors for different approaches [9]. However, there is a need for more internal replication studies from different fields, and more syntheses of these studies given the literature that exists already—see Shadish et al. [10] or, for an example in health systems, Fretheim et al. [11].

On the specific point by Saldanha et al. about justifying inclusion of nonrandomized studies because of rare outcomes, it is true that in many areas there may not be studies that are primarily conducted to measure rare outcomes, which would necessitate the production of QEDs and their incorporation in evidence synthesis. We also note that, if the RCT literature exists which reports underpowered outcomes, these can be statistically combined in meta-analysis. A protocol for one such study, which aims to incorporate many underpowered RCTs as well as available QEDs on the impacts of WASH on mortality in childhood, was recently published in *Campbell Systematic Reviews* [12]. This argument extends beyond estimation of the pooled effect to the often more interesting questions surrounding

heterogeneity. Pooling large numbers of underpowered studies across designs usefully exposes the deficiencies in an evidence base, particularly with respect to inconsistency. It also increases the probability of identifying the drivers of variation where they exist even with highly multidimensional data. However, where outcomes are delayed in onset, pooling RCTs is unlikely to overcome the limitations of individual studies. QEDs are therefore better justified with reference to the measurement of long-term effects, which may be difficult or impossible in RCTs due to concerns about contamination of controls [13].

As Saldanha et al. note, nonrandomized studies need to be appraised carefully and formally in systematic reviews, using risk-of-bias assessment, to separate the “chaff” (that is, studies that do not meet minimum requirements of causal inference) from the “wheat” (studies that do). We are developing a risk-of-bias tool for reviewers working on topics in social policy and related fields like criminology, economics, education, ecology and public health, that evaluates RCTs and QEDs using consistent criteria. However, we also believe that risk-of-bias tools need to be adapted to the specifics of the review that is being conducted, which means that biases important for certain types of contexts or literatures may be less important for others. For example, blinding is likely to be a useful discriminatory risk-of-bias component in reviews of self-reported outcomes like illness or pain, but the difference between spatial and temporal design elements may be more critical in an ecological review.

Incorporating QEDs alongside RCTs in systematic reviews is an operating principle of the Campbell Collaboration. For example, 80% of Campbell systematic reviews incorporate QEDs and this number is increasing. The Campbell Collaboration [14] acknowledges that RCTs may be conducted as research and demonstration projects, thus may have limited generalizability to other settings. The Campbell guidance states: “it is useful for a review to include all methodologically credible evidence about the effects of the intervention so long as the limitations of the different types of research are explicitly recognized.” (p.9).

Our perspective is not to view RCTs and QEDs as categorically distinct but rather as both representing overlapping continuums with respect to their strength of evidence. There are research questions for which high-quality QEDs may provide the best evidence, particularly when the interest is in the effectiveness of an intervention or social policy under real-world conditions with typical, rather than ideal, implementation levels. Stated differently, RCTs may provide a great answer to the wrong question, whereas a QED may provide a good answer to the right question. From this perspective, the decision to include QEDs in a systematic review goes beyond a justification based on an absence of RCTs to an assessment of the fit between the goals of the systematic reviews and the nature and strength of the evidence provided by the available

literature, be it RCTs, QEDs, or a mix of the two. Thus, we fully support the guidance provided by Saldanha et al. but have an even more expansive view of the potential benefits of QEDs for evidence synthesis. In the areas where we commonly work, we would argue that the exclusion of QEDs should be explicitly justified based on arguments about the relevance of the available RCTs to the policy and practice questions the review aims to address.

References

- [1] Saldanha IJ, Adam GP, Bañez LL, Bass EB, Berliner E, Devine B, et al. Inclusion of nonrandomized studies of interventions in systematic reviews of interventions: updated guidance from the agency for Healthcare research and quality effective Healthcare program. *J Clin Epidemiol* 2022. <https://doi.org/10.1016/j.jclinepi.2022.08.015>.
- [2] Sterne JAC, Hernán M, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *Br Med J* 2016;355:i4919.
- [3] Shadish W, Cook T, Campbell D. *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth CENGAGE Learning; 2002.
- [4] National Academies of Sciences, Engineering, and Medicine. *The future of education research at IES: Advancing an equity-oriented science*. Washington, DC: The National Academies Press; 2022.
- [5] Bärnighausen T, Tugwell P, Röttingen JA, Shemilt I, Rockers P, Geldsetzer P, et al. Quasi-experimental study designs series-paper 4: uses and value. *J Clin Epidemiol* 2017;89:21–9.
- [6] Bradford-Hill A. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295–300.
- [7] Chaplin D, Cook T, Zurovac J, Coopersmith J, Finucane M, Vollmer L, et al. The internal and external validity of the regression discontinuity design: a meta-analysis of 15 within-study comparisons: methods for policy analysis. *J Policy Anal Manage* 2018; 37(2):403–29.
- [8] Sharma Waddington H, Villar PF, Valentine JC. Can non-randomised studies of interventions provide unbiased effect estimates? A systematic review of internal replication studies. *Eval Rev* 2022. 193841X221116721.
- [9] Zurovac J, Cook TD, Deke J, Finucane MM, Chaplin DD, Coopersmith JS, et al. Absolute and relative bias in eight common observational study designs: evidence from a meta-analysis. Working paper. 2021. Available at <https://arxiv.org/abs/2111.06941>. Accessed November 25, 2022.
- [10] Shadish WR, Clark MH, Steiner PM. Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *J Am Stat Assoc* 2008;103: 1334–44.
- [11] Fretheim A, Soumerai SB, Zhang F, Oxman AD, Ross-Degnan D. Interrupted time-series analysis yielded an effect estimate concordant with the cluster-randomized controlled trial result. *J Clin Epidemiol* 2013;66:883–7.
- [12] Sharma Waddington H, Cairncross S. PROTOCOL: water, sanitation and hygiene (WASH) for reducing childhood mortality in low- and middle-income countries. *Campbell Syst Rev* 2020;17(1):e1135.
- [13] Welch VA, Ghogomu E, Hossain A, Awasthi S, Bhutta ZA, Cumberbatch C, et al. Mass deworming to improve developmental health and wellbeing of children in low-income and middle-income countries: a systematic review and network meta-analysis. *Lancet Glob Health* 2017;5(1):e40–50.
- [14] Campbell Collaboration. *Campbell systematic reviews: policies and guidelines*. Version 1.8. 2021. Available at https://onlinelibrary.wiley.com/pb-assets/assets/18911803/Campbell%20Policies%20and%20Guidelines%20_May3%202022-1653054593497.pdf. Accessed September 27, 2022.