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 Waddingtona,*, David B. Wilsonb, Terri Pigottc, Gavin Stewarde, Ariel M. Aloe, Peter Tugwellf, Vivian Welchg

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

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).

https://doi.org/10.1016/j.jclinepi.2022.11.005
0895-4356/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
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 nonrandomized 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
and strength of the evidence provided by the available.

tween the goals of the systematic reviews and the nature
QEDs in a systematic review goes beyond a justification
question. From this perspective, the decision to include
RCTs may provide a great answer to the wrong question,
rather than ideal, implementation levels. Stated differently,
or social policy under real-world conditions with typical,
quality QEDs may provide the best evidence, particularly
evidence. There are research questions for which high-

gorically distinct but rather as both representing overlap-
spatial and temporal design elements may be more critical
in an ecological review.

Incorporating QEDs alongside RCTs in systematic re-
views is an operating principle of the Campbell Collab-
oration. 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
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 cate-
gorically distinct but rather as both representing over-
lapping 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 differen-
ly, 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 be-

References

atic reviews of interventions: updated guidance from the agency for
Healthcare research and quality effective Healthcare program. J Clin
[3] Shadish W, Cook T, Campbell D. Experimental and quasi-experi-
mental designs for generalized causal inference. Belmont, CA: Wads-
worth CENGAGE Learning; 2002.
future of education research at IES: Advancing an equity-oriented
[6] Bradford-Hill A. The environment and disease: association or causa-
discontinuity design: a meta-analysis of 15 within-study compar-
studies of interventions provide unbiased effect estimates? A system-
193841X221116721.
observational study designs: evidence from a meta-analysis. Working
November 25, 2022.
[10] Shadish WR, Clark MH, Steiner PM. Can nonrandomized experi-
ments yield accurate answers? A randomized experiment comparing
random and nonrandom assignments. J Am Stat Assoc 2008;103:
1334–44.
terrupted time-series analysis yielded an effect estimate concordant
with the cluster-randomized controlled trial result. J Clin Epidemiol
[12] Sharma Waddington H, Cairncross S. PROTOCOL: water, sanitation
and hygiene (WASH) for reducing childhood mortality in low- and
and wellbeing of children in low-income and middle-income
countries: a systematic review and network meta-analysis. Lancet
[14] Campbell Collaboration. Campbell systematic reviews: policies and
wiley.com/pb-assets/assets/18911803/Campbell%20Guidelines%20May%202022-1653054593497.pdf. Accessed
September 27, 2022.