ELSEVIER

Journal of Clinical Epidemiology ■ (2017) ■

Journal of Clinical Epidemiology

JCE SERIES

Quasi-experimental study designs series—paper 6: risk of bias assessment

Hugh Waddington[a,*], Ariel M. Aloe[b], Betsy Jane Becker[c], Eric W. Djimeu[a], Jorge Garcia Hombrados[d], Peter Tugwell[e], George Wells[e], Barney Reeves[f]

[a]*International Initiative for Impact Evaluation, New Delhi, India*
[b]*University of Iowa, Iowa City, IA, USA*
[c]*Florida State University, Tallahassee, FL, USA*
[d]*University of Sussex, Brighton, UK*
[e]*University of Ottawa, Ottawa, Ontario, Canada*
[f]*University of Bristoln, Bristol, UK*

Accepted 6 February 2017; Published online xxxx

Abstract

Objectives: Rigorous and transparent bias assessment is a core component of high-quality systematic reviews. We assess modifications to existing risk of bias approaches to incorporate rigorous quasi-experimental approaches with selection on unobservables. These are nonrandomized studies using design-based approaches to control for unobservable sources of confounding such as difference studies, instrumental variables, interrupted time series, natural experiments, and regression-discontinuity designs.

Study Design and Setting: We review existing risk of bias tools. Drawing on these tools, we present domains of bias and suggest directions for evaluation questions.

Results: The review suggests that existing risk of bias tools provide, to different degrees, incomplete transparent criteria to assess the validity of these designs. The paper then presents an approach to evaluating the internal validity of quasi-experiments with selection on unobservables.

Conclusion: We conclude that tools for nonrandomized studies of interventions need to be further developed to incorporate evaluation questions for quasi-experiments with selection on unobservables.

Keywords: Risk of bias; Systematic review; Meta-Analysis; Quasi-experiment; Natural experiment; Instrumental variables; Regression discontinuity; Interrupted time series; Difference in differences

1. Introduction

Researchers in health and the social sciences quantify treatment effects—that is, changes in outcomes which are attributed to a particular intervention—using a range of nonrandomized approaches, also called quasi-experiments (QEs) [1–3]. QEs are quantitative studies which are used to make causal inferences when treatment is by definition not randomly assigned. There are two main types of QE study: designs which are able to adjust for unobservable sources of confounding (“selection on unobservables”); and methods which adjust for observables directly (e.g., analysis of variance or adjusted regression analysis) whose validity is based on the assumption of unconfoundedness [4,5]. In this paper, we discuss explicitly approaches to control for selection on unobservables, including difference in differences (DID), instrumental variables, interrupted time series (ITS), natural experiments, and regression discontinuity designs. Often these designs are combined with methods to control for observable confounding such as statistical matching [e.g., propensity score matching (PSM)].

All quantitative causal studies are subject to biases relating to design (internal validity) and methods of statistical analysis (statistical conclusion validity) [3]. In the same way that experimental studies [randomized controlled trials (RCTs) can have methodological problems in implementation (e.g., contamination of controls, poor allocation concealment, nonrandom attrition, and so on), inappropriately designed or executed QEs will not generate good causal evidence. QE studies are, however, potentially at higher risk of bias than their experimental counterparts [6,7], with perhaps the most critical biases for causal

* Corresponding author. Tel.: +44-7779-261108; fax: +44-2030-738303.
E-mail address: hwaddington@3ieimpact.org (H. Waddington).

http://dx.doi.org/10.1016/j.jclinepi.2017.02.015
0895-4356/

What is new?

Key findings

- Rigorous nonrandomized studies use design-based approaches which can control for unobservable sources of confounding. These include difference studies, instrumental variables estimation, interrupted time series, natural experiments, and regression discontinuity designs. Systematic critical appraisal of these studies requires identification of the design and assessment of the methodology, which existing risk of bias tools can incorporate.

What this adds to what was known?

- A review of risk of bias tools suggests that they provide, to different degrees, incomplete transparent criteria to assess rigorous nonrandomized studies. We assess modifications to existing approaches to assess bias, based on study design and methods of analysis.

What is the implication and what should change now?

- Current tools used to assess bias in systematic reviews can be modified to incorporate specific evaluation questions to assess nonrandomized studies with selection on unobservables. Work is underway to incorporate these approaches into Cochrane's risk of bias tool in nonrandomized studies of interventions.

inference being confounding and bias in selection of the reported result. In addition, the assessment of QEs requires greater qualitative appraisal of potential biases than RCTs, which in many cases may need to draw on advanced theoretical and statistical knowledge [4]. At the same time, QEs typically have a number of distinct advantages over experiments because they do not interfere in the natural data generation process [8].

Systematic critical appraisal, operationalized through ''risk of bias'' assessment, gives assurance of the credibility of the point estimates provided causal studies [9] and their trustworthiness for decision making [10]. Risk of bias tools provide transparency about the judgments made by reviewers when performing assessments. They are usually organized around particular domains of bias and provide specific ''signaling questions'' which enable reviewers to evaluate the likelihood of bias.

This paper discusses how to operationalize risk of bias assessment for QEs with selection on unobservables. A glossary of technical terms used is provided in the Appendix at www.jclinepi.com. Section 2 discusses internal validity, and Section 3 reviews existing risk of bias tools. Section 4 presents proposed evaluation criteria. Section 5 proposes an agenda for research in the further development of a risk of bias tool. Section 6 concludes.

2. Internal validity of QEs

Habicht et al. [11] distinguish probability evaluation designs, which are able to quantify with statistical precision the change in outcomes attributed to a treatment, from plausibility designs, which attempt to rule out observable confounding through use of a comparison group but are unable to address important sources of bias, in particular those arising from unobservables. These authors explicitly limit probability evaluations to RCTs. However, evidence is emerging which suggests QEs which use credible methods to address unobservable confounding can produce the same effect sizes as RCTs in pooled analysis (Table 1). We note that authors and journal editors may have incentives for selective publishing of favorable comparisons between randomized and nonrandomized studies. The examples presented in Table 1 are from systematic reviews (SRs) of socioeconomic interventions in low- and middle-income countries supported by the Campbell Collaboration International Development Coordinating Group (IDCG). The findings on experimental and quasi-experimental approaches are representative of the body of evidence in SRs supported by the IDCG. Other examples of comparisons of RCTs and QEs include Lipsey and Wilson [15] who provide a meta-analysis of North American social programs and Vist et al. [16] who compare RCTs and cohort studies in health care studies. Evidence is also available from (within-study) design replication—that is, studies which attempt to compare the same experimental treatment groups with nonrandomized comparison groups using quasi-experimental methods. One meta-study suggested significant differences between results from RCTs and QEs for US and European labor market programs [17]. However, design replications using well-conducted quasi-experimental methods, in which participation has been carefully modeled, have also shown the same results as the RCTs they are replicating [18,19].

As noted by Duvendack et al. [20], effect sizes estimated from nonrandomized studies may differ empirically from those from RCTs due to differences in the population sampled and the type of treatment effect estimated.

QEs modeling selection on unobservables account for confounding by design, either through knowledge about the method of allocation or in the methods of analysis used. They are considered more credible in theory than approaches based on unconfoundedness which rely solely on observable covariate adjustment [21,5,3].

In QE designs that use information about the allocation process to estimate a treatment effect, the ability of the