Health Policy 119 (2015) 511–521

Contents lists available at ScienceDirect

Health Policy

journal homepage: www.elsevier.com/locate/healthpol

Inclusion of quasi-experimental studies in systematic reviews of health systems research

Peter C. Rockers[a,b], John-Arne Røttingen[c,d,e,f], Ian Shemilt[g], Peter Tugwell[h,i], Till Bärnighausen[d,j,*]

[a] Department of Global Health, Boston University School of Public Health, Boston, USA
[b] Center for Global Health and Development, Boston University, Boston, USA
[c] Norwegian Institute of Public Health, Oslo, Norway
[d] Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, USA
[e] Department of Health Management and Health Economics, Institute for Health and Society, University of Oslo, Norway
[f] Harvard Global Health Institute, Harvard University, Cambridge, USA
[g] Behavior and Health Research Unit, University of Cambridge, Cambridge, UK
[h] Department of Medicine, University of Ottawa, Canada
[i] Department and Epidemiology and Community Health, University of Ottawa, Canada
[j] Wellcome Trust Africa Centre for Health and Population Studies, University of KwaZulu-Natal, Mtubatuba, South Africa

ARTICLE INFO

Article history:
Received 20 April 2014
Received in revised form
29 September 2014
Accepted 11 October 2014

Keywords:
Quasi-experimental studies
Systematic reviews
Regression discontinuity designs
Instrumental variable analysis
Interrupted time series
Difference-in-difference studies

ABSTRACT

Systematic reviews of health systems research commonly limit studies for evidence synthesis to randomized controlled trials. However, well-conducted quasi-experimental studies can provide strong evidence for causal inference. With this article, we aim to stimulate and inform discussions on including quasi-experiments in systematic reviews of health systems research. We define quasi-experimental studies as those that estimate causal effect sizes using exogenous variation in the exposure of interest that is not directly controlled by the researcher. We incorporate this definition into a non-hierarchical three-class taxonomy of study designs – experiments, quasi-experiments, and non-experiments. Based on a review of practice in three disciplines related to health systems research (epidemiology, economics, and political science), we discuss five commonly used study designs that fit our definition of quasi-experiments: natural experiments, instrumental variable analyses, regression discontinuity analyses, interrupted times series studies, and difference studies including controlled before-and-after designs, difference-in-difference designs and fixed effects analyses of panel data. We further review current practices regarding quasi-experimental studies in three non-health fields that utilize systematic reviews (education, development, and environment studies) to inform the design of approaches for synthesizing quasi-experimental evidence in health systems research. Ultimately, the aim of any review is practical: to provide useful information for policymakers, practitioners, and researchers. Future work should focus on building a consensus among users and producers of systematic reviews regarding the inclusion of quasi-experiments.

1. Introduction

Systematic reviews of health systems research can provide policymakers with information that can be used to formulate health policies based on the best available

* Corresponding author at: Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, USA.
Tel.: +1 6173790372.
E-mail address: tbaernig@hsph.harvard.edu (T. Bärnighausen).

http://dx.doi.org/10.1016/j.healthpol.2014.10.006
0168-8510/

evidence [1]. Systematic reviews have recently been conducted on a wide variety of health systems topics, including health care provider payment systems [2–4], demand-side incentives for health care use [5,6], and clinical task shifting [7–9]. In general, systematic reviews aim to synthesize in an objective and reproducible way the full body of evidence available on a given topic [10,11]. In addition to synthesizing evidence, systematic reviews typically also incorporate assessments of the quality of that evidence [12,13]. Authors of systematic reviews, as experts in their fields, add value to the primary research they synthesize by delineating between evidence of differing quality based on considerations of risk of bias, precision, consistency and relevance. This task is difficult because relevant evidence is not necessarily of high-quality; authors must carefully consider potential trade-offs between relevance and different dimensions of research quality when deciding which studies to include in systematic reviews.

Authors of systematic reviews frequently apply study design eligibility criteria when selecting studies for inclusion and this may often reflect explicit or implicit judgments about which evidence is of sufficient quality to inform policy. In a previous meta-review of systematic reviews of studies investigating the effectiveness of health systems interventions, Rockers et al. [14] found that half of all reviews limited their study design inclusion criteria to randomized controlled trials, which are only rarely possible in health systems research [15], while most of the other half of systematic reviews limited evidence synthesis to primary studies using a subset of four designs suggested by the Effective Practice and Organization of Care (EPOC) Review Group of The Cochrane Collaboration: randomized controlled trials (RCT), quasi-randomized controlled trials (QRCT), controlled before-and-after (CBA) studies, and interrupted time series (ITS) studies [10,14]. CBA and ITS studies have been classified by the EPOC Review Group as quasi-experimental (QE) [16]. Other designs that may similarly be classified as QE are not included in the EPOC set and have been excluded from most systematic reviews of health systems research.

The term "quasi-experimental" was first introduced by Campbell and Stanley [17] and has since been defined and employed by several authors from diverse disciplines (see Table 1). The wording of the definitions given in Table 1 varies; however, all authors identify similar features as essential to QE studies. Most importantly, all recognize QE studies as those that identify and use for causal effect size estimation exogenous variation in the exposure of interest despite lack of researcher control over the process assigning the exposure. Exogenous variation refers to variation determined outside the system of causal relationships under study [18].

However, each author's attempt to enumerate the universe of study designs that should be classified as QE has arrived at different conclusions [19–22]. No definitive list of QE study designs currently exists.

Delineating between QE studies and other non-RCTs does not imply that QE studies are necessarily of higher quality, just as RCTs are not necessarily of higher quality than non-RCTs. All study designs have identifying assumptions that must be validated before causal inference can be drawn, and all studies to be included in a systematic review must be assessed for quality according to both general principles of unbiased causal effect estimation and design-specific risk of bias standards [23]. Given the validity of certain assumptions, even the simplest of non-randomized studies can provide valid causal effect estimates and *ex ante* exclusion of studies from systematic reviews based on design may reduce the value of evidence synthesis. In many respects, the application of inclusion criteria based on study design characteristics is not a necessity, but rather can serve the practical purpose of relieving review authors of the responsibility of judging the quality of certain types of studies for which either insufficient information is provided to validate assumptions or for which the review authors themselves have insufficient expertise. For example, for relationships estimated using simple regression models, it may be impossible to assess the risk of unobserved confounding if too little is known regarding potential confounders in a specific context.

While the theoretical underpinnings of our taxonomy lead us to conclude that a hierarchy of designs is not appropriate, we concede that, in practice, the assumptions needed to argue that effect estimates are causal are often easier to verify for particular study designs as compared to others. In particular, it is often easier to verify that the exposure of interest was randomly allocated in an experiment, where the randomization process is controlled by the researcher, than it is to verify the same assumption in a natural experiment. For "strong" quasi-experimental designs the assumptions required for valid causal inference are weaker, and thus easier to verify, than those for "weak" designs. (We develop the idea of "strong" and "weak" quasi-experimental studies in more detail when discussing specific study designs in Section 3.) Study design inclusion criteria may also lessen the need for subjective judgments on the part of authors, increasing the systematic nature of reviews. There are ongoing debates regarding whether use of study design inclusion criteria are appropriate for systematic reviews and other forms of evidence synthesis [24,25]. However these debates will be resolved, as a general principle it is important in systematic reviews to clearly describe the study inclusion criteria, the approaches taken to rate quality of primary studies, and the study quality ratings themselves, so that readers can judge the merits of a particular systematic review and the recommendations derived from it.

In this paper, we aim to inform discussions on the inclusion of QE study designs in systematic reviews of health systems research. First, we outline a conceptual framework for understanding this issue. We present our definition of "quasi-experimental" and develop a related taxonomy in order to clarify how studies that fit this definition may differ from other studies in ways that are relevant for authors of systematic reviews as well as policy makers. Second, we identify a set of QE study designs that fit our definition and that are commonly employed in three disciplines: epidemiology, economics, and political science. We focus on these particular disciplines because we believe they capture the range of methods that are found in policy-relevant health systems research. We attempt to clarify any overlap across disciplines that may be obfuscated