Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

ANNUAL REVIEWS

Annual Review of Public Health

Selecting and Improving Quasi-Experimental Designs in Effectiveness and Implementation Research

Margaret A. Handley,[1,2,3] Courtney R. Lyles,[2,3] Charles McCulloch,[1] and Adithya Cattamanchi[3,4]

[1]Department of Epidemiology and Biostatistics, University of California, San Francisco, California 94110, USA; email: margaret.handley@ucsf.edu

[2]Department of Medicine, Division of General Internal Medicine at Zuckerberg San Francisco General Hospital and Trauma Center, University of California, San Francisco, California 94110, USA

[3]Center for Vulnerable Populations at Zuckerberg San Francisco General Hospital and Trauma Center, University of California, San Francisco, California 94110, USA

[4]Department of Medicine, Division of Pulmonary and Critical Care Medicine at Zuckerberg San Francisco General Hospital and Trauma Center, University of California, San Francisco, California 94110, USA

ANNUAL REVIEWS **Further**

Click here to view this article's online features:

- Download figures as PPT slides
- Navigate linked references
- Download citations
- Explore related articles
- Search keywords

Annu. Rev. Public Health 2018. 39:5–25

First published as a Review in Advance on January 12, 2018

The *Annual Review of Public Health* is online at publhealth.annualreviews.org

https://doi.org/10.1146/annurev-publhealth-040617-014128

Copyright © 2018 Margaret A. Handley et al. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See credit lines of images or other third-party material in this article for license information

This article is part of a symposium on Implementation Science and Public Health. For a list of other articles in this symposium, see **http://www.annualreviews.org/toc/publhealth/39/1**

Keywords

quasi-experimental design, stepped wedge, interrupted time series, prepost, implementation science, external validity

Abstract

Interventional researchers face many design challenges when assessing intervention implementation in real-world settings. Intervention implementation requires holding fast on internal validity needs while incorporating external validity considerations (such as uptake by diverse subpopulations, acceptability, cost, and sustainability). Quasi-experimental designs (QEDs) are increasingly employed to achieve a balance between internal and external validity. Although these designs are often referred to and summarized in terms of logistical benefits, there is still uncertainty about (*a*) selecting from among various QEDs and (*b*) developing strategies to strengthen the internal and external validity of QEDs. We focus here on commonly used

OPEN ACCESS CC BY

QEDs (prepost designs with nonequivalent control groups, interrupted time series, and stepped-wedge designs) and discuss several variants that maximize internal and external validity at the design, execution and implementation, and analysis stages.

INTRODUCTION

Public health practice involves implementing or adapting evidence-based interventions into new settings to improve health for individuals and populations. Such interventions typically include one or more of the 7 Ps (programs, practices, principles, procedures, products, pills, and policies) (7). Both public health and clinical research have increasingly sought to generate practice-based evidence on a wide range of interventions, which in turn has led to a greater focus on developing intervention research designs that can be applied in real-world settings (2, 7–9, 19, 24, 25).

Randomized controlled trials (RCTs) in which individuals are assigned to intervention or control (standard-of-care or placebo) arms are considered the gold standard for assessing causality and, as such, are a first choice for most intervention research. Random allocation minimizes selection bias and maximizes the likelihood that measured and unmeasured confounding variables are distributed equally, enabling any difference in outcomes between the intervention and control arms to be attributed to the intervention under study. RCTs can also involve random assignment of groups (e.g., clinics, worksites, or communities) to intervention and control arms, but a large number of groups are required to realize the full benefits of randomization. Traditional RCTs strongly prioritize internal validity over external validity by employing strict eligibility criteria and rigorous data collection methods.

Alternative research methods are needed to test interventions for their effectiveness in many real-world settings and, when evidence-based interventions are known, to prepare for spreading or scaling up these interventions to new settings and populations (22, 41). In real-world settings, random allocation of the intervention may not be possible or fully under the control of investigators because of practical, ethical, social, or logistical constraints. For example, when partnering with communities or organizations to deliver a public health intervention, it may not be acceptable that only half of individuals or sites receive an intervention. As well, the timing of intervention rollout may be determined by an external process outside the control of the investigator, such as a mandated policy. Also, when self-selected groups are expected to participate in a program as part of routine care, ethical concerns associated with random assignment would arise, for example, the withholding or delaying of a potentially effective treatment or the provision of a less effective treatment for one group of participants (51). As described by Peters et al., "[I]mplementation research seeks to understand and work within real-world conditions, rather than trying to control for these conditions or to remove their influence as causal effects" (41). For all these reasons, a blending of the design components of clinical effectiveness RCTs and implementation research is feasible and desirable, and this review covers both. Such blending of effectiveness and implementation components within a study can provide benefits beyond either research approach alone (17)—for example, by leading to faster uptake of interventions by simultaneously testing implementation strategies.

Because assessment of intervention effectiveness and implementation in real-world settings requires increased focus on external validity (including the consideration of factors enhancing intervention uptake by diverse subpopulations, acceptability to a wide range of stakeholders, cost, and sustainability) (33), investigators need interventional research designs that are more relevant to

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

TERMS AND DEFINITIONS

External validity: describes the extent to which a research conclusion can be generalized to the population or to other settings

Internal validity: refers to the extent to which a study is capable of establishing causality; it is related to the degree it minimizes error or bias

Interrupted time series (ITS) design: multiple observations are evaluated for several consecutive points in time before and after intervention within the same individual or group

Inverse rollout: sites are rolled out to receive the intervention using a structured approach to create balance between the sites over the rollout time period, using a sample characteristic that is ordered (and then reverse ordered). Often size or geography is used (e.g., 1,2,3,4 for size followed by 4,3,2,1)

Nonequivalent control group: a control group that is not randomly assigned to receive or not receive the intervention. An intact group is selected that is thought to be similar to the intervention group

Partial randomization: a type of stratified randomization, with strata constructed for potential confounding variables; randomization occurs separately within each stratum

Prepost design: a QED with data collected before and after an intervention is introduced and then compared. A control group can be added for a prepost design with a nonequivalent control group

Quasi-experimental design (QED): includes a wide range of nonrandomized or partially randomized prepost intervention studies

Stepped-wedge design (SWD): a type of crossover design where the time of crossover is intentionally or randomly assigned

Washout period: time period during which a prior practice or intervention is stopped and a new one is implemented, for which both interventions may be operating, and thus the data are excluded from the analysis

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

the potential, hoped-for treatment population and that achieve a better balance between internal and external validity, compared with an RCT. Quasi-experimental designs (QEDs), which first gained prominence in social science research (10), are increasingly being employed to fill this need (see the sidebar titled Terms and Definitions).

QEDs test causal hypotheses but, in lieu of fully randomized assignment of the intervention, seek to define a comparison group or time period that reflects the counterfactual (i.e., outcomes if the intervention had not been implemented) (45). QEDs seek to identify a comparison group or time period that is as similar as possible to the treatment group or time period in terms of baseline (preintervention) characteristics. QEDs can include partial randomization, such as in stepped-wedge designs (SWDs) when there is predetermined (and nonrandom) stratification of sites but the order in which sites within each stratum receive the intervention is assigned randomly. For example, strata that are determined by size or perceived ease of implementation may be assigned to receive the intervention first. However, within those strata, the specific sites themselves are randomly selected to receive the intervention across the time intervals specified in the study. In all cases, the key threat to internal validity of QEDs is a lack of similarity between the comparison and intervention groups or time periods, owing to differences in characteristics of the people, sites, or time periods involved.

Previous reviews in the *Annual Review of Public Health* have focused on the importance and use of QEDs and other methods to enhance causal inference when evaluating the impact of an intervention that has already been implemented (4, 7, 8, 16). Design approaches in this case often include creating a post hoc comparison group for a natural experiment or identifying pre- and

COMMON THREATS TO INTERNAL VALIDITY OF QUASI-EXPERIMENTAL DESIGNS EVALUATING INTERVENTIONS IN REAL-WORLD SETTINGS

History bias: events other than the intervention occurring at the same time that may influence the results
Selection bias: systematic differences in subject characteristics between the intervention and control groups that are related to the outcome
Maturation bias: changes among individuals in the groups, differently over time, resulting in effects in addition to (or rather than) the treatment condition that may change the performance of participants in the posttest relative to the pretest
Lack of blinding: awareness of group assignment that influences those delivering or receiving the intervention
Differential dropout: attrition affecting the intervention or the control group differently, resulting in selection bias and/or loss of statistical power
Variability in interactive effects: implementation of an intervention with multiple components that varies across the implementation process and/or by sites
Source: Adapted from Reference 52

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

postintervention data to then conduct an interrupted time series study. Analysis phase approaches often utilize techniques such as prepost, regression adjustment, scores, difference-in-differences, synthetic controls, interrupted time series, regression discontinuity, and instrumental variables (4, 7, 16). Although these articles summarize key components of QEDs [e.g., interrupted time series (ITS)], as well as analysis-focused strategies (regression adjustment, propensity scores, difference-in-differences, synthetic controls, and instrumental variables), there is still uncertainty about (*a*) selecting from among various QEDs in the preimplementation design phase and (*b*) developing strategies to strengthen internal and external validity before and during the implementation phase.

In this article, we discuss the a priori choice of a QED when evaluating the impact of an intervention or policy for which the investigator has some element of design control related to (*a*) the ordering of intervention allocation (including random and nonrandom approaches); (*b*) the selection of sites or individuals; and/or (*c*) the timing and frequency of data collection. In the next section, we discuss the main QEDs used for prospective evaluations of interventions in real-world settings and their advantages and disadvantages with respect to addressing threats to internal validity (see the sidebar titled Common Threats to Internal Validity of Quasi-Experimental Designs Evaluating Interventions in Real-World Settings). Following this summary, we discuss opportunities to strengthen their internal validity, illustrated with examples from the literature. Then we propose a framework for key decision points that lead to different QED options. We conclude with a brief discussion of incorporating additional design elements to capture the full range of relevant implementation outcomes in order to maximize external validity.

QUASI-EXPERIMENTAL DESIGNS FOR PROSPECTIVE EVALUTION OF INTERVENTIONS

Table 1 summarizes the main QEDs that have been used for prospective evaluation of health interventions in real-world settings: prepost designs with a nonequivalent control group, ITS, and SWDs. We do not include prepost designs without a control group in this review as, in general, QEDs are primarily those designs that identify a comparison group or time period that is as similar

Table 1 Overview of commonly used QED in intervention research

QED design	Key design elements	Advantages	Disadvantages
Prepost with nonequivalent control group	Those receiving the intervention are compared with those not receiving it Analysis is usually based on estimating the difference in the amount of change over time in the outcome of interest between the two groups, beginning with the intervention and moving forward in time The two groups can also be examined from the same population using before-and-after intervention cohorts	Simplicity of data collection when using a small number of time points Associated lower cost Less cumbersome to implement than other designs	Temporal biases are a substantial risk and may result in regression to the mean or overinterpretation of intervention effects Quality of data may vary in different time periods, resulting in measurement error Nonequivalent sites may not be comparable for important covariates
Interrupted time series	Multiple observations are assessed for a number of consecutive points in time before and after the intervention within the same individual or group	Useful for when there is a small number of communities or groups, as each group acts as its own control May be the only option for studying the impacts of large-scale health policies	Requires a large number of measurements May not be feasible for geographically dispersed areas
Stepped-wedge design	Intervention is rolled out over time, usually at the site level Participants who initially do not receive the intervention later cross over to receive the intervention; those who wait provide control data during the time when others receive the intervention, reducing the risk of bias due to time and time-dependent covariates The study can be based on serial cross-sectional data collected by sites for different time periods (sites cross over) or by following a cohort of the same individuals over time (individuals cross over)	All clusters or wait list groups eventually receive the intervention Investigators do not need to supply the intervention in all sites in a short time frame, using staggered implementation	May not be able to randomly assign rollout of sites, thereby potentially jeopardizing internal validity Cannot guarantee that everyone in each cluster or list will receive the intervention during the time that the cluster is receiving the intervention Often takes longer than other designs to implement Control data must be collected or ascertained from sites or participants Site differences and implementation processes can vary significantly over time Risk of contamination in later sites or intervention fatigue—both can wash out potential intervention effects

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

as possible to the treatment group or time period in terms of baseline (preintervention) characteristics (52). Below, we describe features of each QED, considering strengths and limitations and providing examples of their use.

Prepost with Nonequivalent Control Group

The first type of QED highlighted in this review is perhaps the most straightforward type of intervention design: the prepost comparison study with a nonequivalent control group (**Figure 1**). In this design, the intervention is introduced at a single point in time to one or more sites, for which there is also a pretest and posttest evaluation period. The prepost differences between these

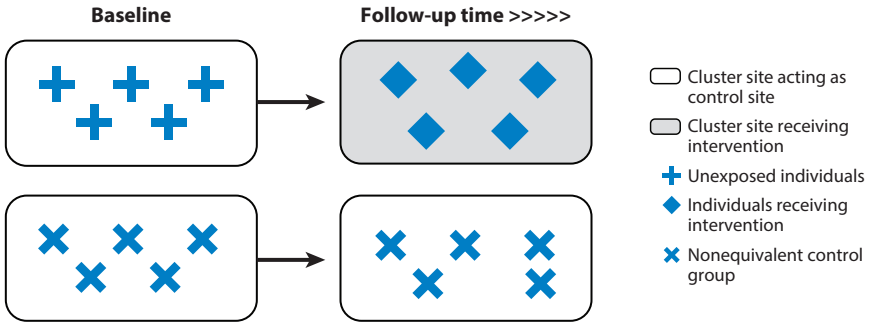

Figure 1

Illustration of the prepost nonequivalent control group design.

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

two sites are then compared. In practice, interventions using this design are often delivered at a higher level, such as to entire communities or organizations.[1] In this design, the investigators identify additional site(s) that are similar to the intervention site to serve as a comparison/control group. However, these control sites are different in some way than the intervention site(s) and thus the term nonequivalent is important in clarifying that there are inherent differences in the treatment and control groups (13).

The strengths of prepost designs are based mainly in their simplicity, such that data are collected usually only at a few points (although sometimes more). However, prepost designs can be affected by several of the threats to internal validity of QEDs presented here. The largest challenges are related to (*a*) history bias, in which events unrelated to the intervention occur (also referred to as secular trends) before or during the intervention period and have an effect on the outcome (either positive or negative) that is not related to the intervention (39); and (*b*) differences between the intervention and control sites because the nonequivalent control groups are likely to differ from the intervention sites in a number of meaningful ways that impact the outcome of interest and can bias results (selection bias).

At this design stage, the first step at improving internal validity would be focused on the selection of a nonequivalent control group(s) for which some balance in the distribution of known risk factors is established. This step can be challenging because there may not be adequate information available to determine how equivalent the comparison group is regarding relevant covariates.

It can be useful to obtain pretest data or baseline characteristics to improve the comparability of the two groups. In the most controlled situations within this design, the investigators might include elements of randomization or matching for individuals in the intervention or comparison site to attempt to balance the covariate distribution. Implicit in this approach is the assumption that the greater the similarity between groups, the smaller the likelihood that confounding will threaten inferences of causality of effect for the intervention (32, 49). Thus, it is important to select this group or multiple groups with as much specificity as possible.

To enhance the causal inference for prepost designs with nonequivalent control groups, the most effective strategies improve the comparability of the control group with regard to potential covariates related to the outcome of interest but are not under investigation. One strategy involves creating a cohort and then using targeted sampling to inform matching of individuals within the

[1]It is important to note that if such randomization would be possible at the site level based on similar sites, then a cluster randomized controlled trial would be an option.

cohort. Matching can be based on demographic and other important factors (e.g., measures of health care access or time period). This design in essence creates a matched, nested case-control design.

Collection of additional data once sites are selected cannot in itself reduce bias but can inform the examination of the association of interest and provide data supporting interpretation that is consistent with the reduced likelihood of bias. These data collection strategies include (*a*) extra data collection points at additional pre- or post- time points (to get closer to an interrupted time series design in effect and examine potential threats of maturation and history bias) and (*b*) collection of data on other dependent variables with a priori assessment of how they will react with time-dependent variables. A detailed analysis can then provide information on the potential effects on the outcome of interest (to understand potential underlying threats due to history bias).

Additionally, some analytic strategies can improve the interpretation of this design to be able to investigate a potential dose–response effect; such strategies include (*a*) analysis for multiple nonequivalent control groups, to determine if the intervention effects are robust across different conditions or settings (e.g., using sensitivity analysis), (*b*) examination within a smaller critical window of the study in which the intervention would be plausibly expected to make the most impact, and (*c*) identification of subgroups of individuals within the intervention community who are known to have received high versus low exposure to the intervention to be able to investigate a potential dose–response effect. **Table 2** provides examples of studies using the prepost nonequivalent control group designs that have employed one or more of these improvement approaches to increase the study's internal validity.

Cousins et al. utilized a nonequivalent control selection strategy to leverage a recent cross-sectional survey among six universities in New Zealand regarding drinking among college-age students (14). In the original survey, there were six sites, and for the control group, five were selected to provide nonequivalent control group data for the one intervention campus. The campus intervention targeted young-adult drinking-related problems and other outcomes, such as aggressive behavior, using an environmental intervention with a community liaison and a campus security program (also known as a Campus Watch program). The original cross-sectional survey was administered nationally to students using a web-based format and was repeated in the years soon after the Campus Watch intervention was implemented in one site. Benefits of the design to support inference include a consistent sampling frame at each control site, such that sites could be combined and also evaluated separately, as well as additional data collection on alcohol sales and consumption over the study period. In a study by Wertz et al. (50), a nonequivalent control group was created using matching for those who were eligible for a health-coaching program and those who opted out of the program (to be compared with those who opted in) among insured patients with diabetes and/or hypertension. Matching was based on propensity scores among those patients according to demographic and socioeconomic factors and medical center location, and a longitudinal cohort was created prior to the intervention [see Basu et al. 2017 (4) for more on this approach].

In the prepost malaria-prevention intervention example from The Gambia, the investigators were studying the introduction of bed nets treated with insecticide on The Gambia's malaria rates, and they collected additional data to evaluate the internal validity assumptions within their design (1). In this study, the investigators introduced bed nets at the village level, using communities not receiving the bed nets as control sites. To strengthen the internal validity, they collected additional data that enabled them to (*a*) determine whether the reduction in malaria rates were most pronounced during the rainy season within the intervention communities, as this was a biologically plausible exposure period in which they could expect the largest effect size difference between intervention and control sites, and (*b*) examine use patterns for the bed nets, on the basis

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

Table 2 Improving quasi-experimental designs, with considerations of internal and external validity

Study/general design	Intervention	Design strategy to improve internal validity	Design strategy to improve external validity
Prepost designs with nonequivalent control group			
Cousins et al. 2014 (14)	Campus Watch program targeting problem drinking and violence at universities in New Zealand 1 university campus with an active intervention, compared with 5 control campuses	**Consistent sampling across and within control sites:** ■ Standardization of independent repeat sampling, survey, and follow-up methods across all sites **Control groups:** ■ 5 sites as control sites studied aggregately and individually **Collection of additional data to support inference:** ■ Consumption and alcohol-related harms data from national surveys to compare data trends over time	Oversampling of indigenous groups to extend the interpretation of findings
Wertz et al. 2012 (50)	Chronic disease management program with pharmacist-based patient coaching within a health care insurance plan in Cincinnati, United States 607 patients in intervention groups, compared to 557 matched control patients	**Prospective cohort sampling to improve control group similarity:** ■ Matching of participants with nonparticipants on demographic and health care access measures (using propensity score matching)	Not applicable
Alonso et al. 1993 (1)	Bed net intervention in The Gambia 41 intervention sites compared with control villages	**Collection of additional data to support inferences:** ■ Examination of data trends during the highest infection times of the year (i.e., rainy season versus dry season) **Subgroup analysis:** ■ Detailed study of those using bed nets within intervention villages (i.e., guaranteed exposure dose to examine dose–response in intervention arm)	Not applicable
Interrupted time series			
Pellegrin et al. 2017 (40)	Hospital discharge program to connect patients to community-based pharmacist follow-up among hospitals in Hawaii 6 intervention and 5 control sites	**Extended period of data collection:** ■ Long baseline period (12 preintervention data points) **Control group inverse rollout/covariate balance:** ■ Intervention rollout staggered on the basis of staff availability (site 1 had 8 postintervention data points, whereas site 8 had two)	Detailed implementation-related process measures monitored (and provided to individual community-based pharmacists regarding their performance) over the entire study period
Robinson et al. 2015 (44)	Hospital discharge program with nurse telephone follow-up and referral among hospitals in Auckland, New Zealand No control group	**Additional analysis:** ■ Examined regression discontinuity during the intervention period to determine if the risk score used to determine eligibility for the program also influenced the outcome	Measured implementation outcomes of whether the intervention was delivered with high fidelity to the protocol

(*Continued*)

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

Table 2 (*Continued*)

Study/general design	Intervention	Design strategy to improve internal validity	Design strategy to improve external validity
Interrupted time series			
Zombré et al. 2017 (54)	Program to remove health care co-payments for high-risk women and children in Burkina Faso 28 intervention health centers compared with 40 control health centers	**Extended period of data collection:** Built into a pilot to collect control data and then extended this work to include additional districts (one intervention, and one nonintervention district) along with 6 additional years of observation **Control group**	Examined sustainability over 72 months of follow-up and associations with clinic characteristics, such as density of workforce
Stepped-wedge design			
Killam et al. 2010 (30)	Rapid antiretroviral treatment intervention among women with HIV in Zambia 8 public-sector clinics	**Site matching:** ■ Matching of 8 sites into four pairs on the basis of the number of pregnant women with HIV expected at each site **Inverse rollout/covariate balance:** ■ The intervention rollout done for one member of the least busy pair, one member of the third busiest pair, one member of the second busiest pair, and one member of the busiest pair; rollout to the remaining pairs proceeded in reverse order **Transition cohort:** ■ A transition cohort during a washout period was established that was later excluded from the analysis, including women who were identified as eligible in the control period of time close to when the intervention was starting	Not applicable
Morrison et al. 2015 (37) See also Dainty et al. 2011 (18)	Program to evaluate out-of-hospital cardiac arrest temperature management protocols in Ontario, Canada 6 regional emergency management systems and 32 hospitals	**Cluster randomization by size:** ■ Randomization at the level of the hospital rather than at the patient level to minimize contamination ■ Hospitals stratified by number of intensive care unit beds (<10 beds versus ≥10 beds as a proxy for hospital size); randomization was done within strata **Transition cohort:** ■ Formalized a transition cohort to test a more passive (i.e., education and protocol guidance) rather than active (i.e., personalized audit and feedback for specific sites) intervention strategy ■ Allowed more time for sites to adopt all elements of the complex intervention before crossing over to the active intervention group	Characterization of system and organizational factors that might affect adoption Collection of longitudinal data relevant to implementation processes that could impact interpretation of findings such as academic versus community affiliation and urban versus rural (hospital bed size)

(*Continued*)

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

Table 2 (*Continued*)

Study/general design	Intervention	Design strategy to improve internal validity	Design strategy to improve external validity
Stepped-wedge design			
Cissé et al. 2016 (11)	Seasonal malaria prophylaxis for children up to age 10 in central Senegal 54 health post catchment areas	**Constrained randomization by geographic indicators and time period**: ■ Constrained randomization of program rollout across regions ■ More sites received the intervention at later stages ($n = 18$) than in the beginning ($n = 9$) ■ Balanced settings for potential confounders, such as distance from river, distance from health center, population size, number of villages, and assessment of ability to implement **Control Group:** ■ Included nine clinics as control sites throughout the study period	Characterization of factors that might affect usage and adherence with longitudinal data Independent evaluations of malaria prophylaxis usage, adherence, and acceptance included prospectively, using routine health cards at the family level and with external assessments from community surveys In-depth interviews conducted across community levels to understand acceptability and other responses to the intervention Embedded study broadened inclusion criteria to focus on a wider age group of at-risk children
Grant et al. 2005 (23)	Mine employees with HIV receiving preventive medication therapy for tuberculosis in South Africa 1,655 employees received treatment at a randomly selected time	**Wait-list individual randomization:** ■ Employees invited from a wait-list in random sequence to attend the workplace HIV clinic	Enumeration of at-risk cohort and estimation of spillover effect beyond those enrolled Using the master enrollment list, estimation of the effect of the intervention among the entire eligible population, not just those enrolled in the intervention over the study period
Ratanawongsa et al. (43); Handley et al. 2011 (26)	A health–information technology (IT) enabled telephone coaching program for patients with diabetes in San Francisco, California, USA 362 patients from 4 clinics received the program using a random wait-list approach	**Wait-list individual randomization:** ■ Patients identified from an actively maintained diabetes registry randomized to receive the coaching intervention immediately or after 6 months **Constrained randomization by language:** ■ Patients randomized to balance enrollment for English, Cantonese, and Spanish speakers over the study period	Use of a framework to assess external validity–related measures for acceptability among patients as well as fidelity measures

(*Continued*)

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

Table 2 (*Continued*)

Study/general design	Intervention	Design strategy to improve internal validity	Design strategy to improve external validity
Stepped-wedge design			
Bailet et al. 2013 (3)	Literacy intervention for prekindergarten children at risk for reading failure in a southern US city 374 children at 102 child care and preschool sites	**Constrained randomization by size then cluster randomization within strata:** ■ Random assignment of clusters (schools) for the timing of the intervention ■ Did not randomize at the site level for large sites; instead, split the schools into 2 groups and randomized at the child level **Site matching:** ■ Matched pairs of child care centers by zip code and percentage of children receiving a state-sponsored financial subsidy; used random assignment within these groups to receive either immediate or deferred enrollment into the intervention	Addition of teacher training in year 2–3 to enhance and measure fidelity Measured completion of each week of the curriculum to enhance assessment of a potential dose–response Refined/adapted intervention in years 2–3, based on initial data
Fernald et al. 2008 (20)	Cash transfer program to mandate child school attendance and receipt of preventive care at the family level in 7 states in Mexico 320 early-intervention and 186 late-intervention (approximately one year later) communities	**Constrained randomization by time period**: ■ More communities randomized to an early-intervention period	Not applicable

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

of how much insecticide was present in the bed nets over time (after regular washing occurred), which aided in calculating a dose–response effect of exposure to the bed net among a subsample of individuals in the intervention community.

Interrupted Time Series

An ITS design involves the collection of outcome data at multiple time points before and after an intervention is introduced at one or more sites (12, 34) (**Figure 2**). The preintervention outcome data are used to establish an underlying trend that is assumed to continue unchanged in the absence of the intervention under study (i.e., the counterfactual scenario). Any change in outcome level or trend from the counterfactual scenario in the postintervention period is then attributed to the impact of the intervention. The most basic ITS design utilizes a regression model that includes only three time-based covariates to estimate the preintervention slope (outcome trend before the intervention), a step or change in level (difference between observed and predicted outcome levels at the first postintervention time point), and a change in slope (difference between post- and preintervention outcome trend) (12, 31).

Whether used for evaluating a natural experiment or, as is the focus here, for prospective evaluation of an intervention, the appropriateness of an ITS design depends on the nature of the intervention and outcome, as well as on the type of data available. An ITS design requires

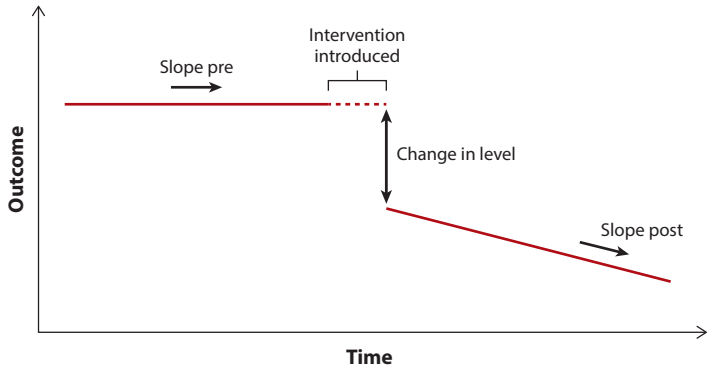

Figure 2

Interrupted time series design.

the pre- and postintervention periods to be clearly differentiated. When used prospectively, the investigator needs to have control over the timing of the intervention. ITS analyses typically involve outcomes that are expected to change soon after an intervention is introduced or after a well-defined lag period. For example, for outcomes such as cancer or incident tuberculosis that develop long after an intervention is introduced and at a variable rate, it is difficult to clearly separate the pre- and postintervention periods. Last, an ITS analysis requires at least three time points in the pre- and postintervention periods to assess trends. In general, a larger number of time points is recommended, particularly when the expected effect size is smaller, data are more similar at time points that are closer together (i.e., autocorrelation), or confounding effects (e.g., seasonality) are present. It is also important for investigators to consider any changes to data collection or recording over time, particularly if such changes are associated with the introduction of the intervention.

In comparison with simple prepost designs in which the average outcome level is compared between the pre- and postintervention periods, the key advantage of ITS designs is that they evaluate for intervention effect while accounting for preintervention trends. Such trends are common owing to factors such as changes in the quality of care, data collection and recording, and population characteristics over time. In addition, ITS designs can increase power by making full use of longitudinal data instead of collapsing all data to single pre- and postintervention time points. The use of longitudinal data can also be helpful for assessing whether intervention effects are short-lived or sustained over time.

Although the basic ITS design has important strengths, the key threat to internal validity is the possibility that factors other than the intervention are affecting the observed changes in outcome level or trend. Changes over time in factors may not be fully accounted for by the preintervention trend. Similarly, the preintervention time period, particularly when short, may not capture seasonal changes in an outcome.

Detailed reviews have been published of variations on the basic ITS design that can be used to enhance causal inference. In particular, the addition of a control group can be particularly useful for assessing the presence of seasonal trends and other potential time-varying confounders (54). Zombré et al. (54) maintained a large number of control sites during the extended study period and were able to evaluate variations in seasonal trends as well as clinic-level characteristics, such as workforce density and sustainability. In addition to including a control group, several analysis-phase strategies can be employed to strengthen causal inference, including adjustment for time-varying confounders and accounting for autocorrelation.

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

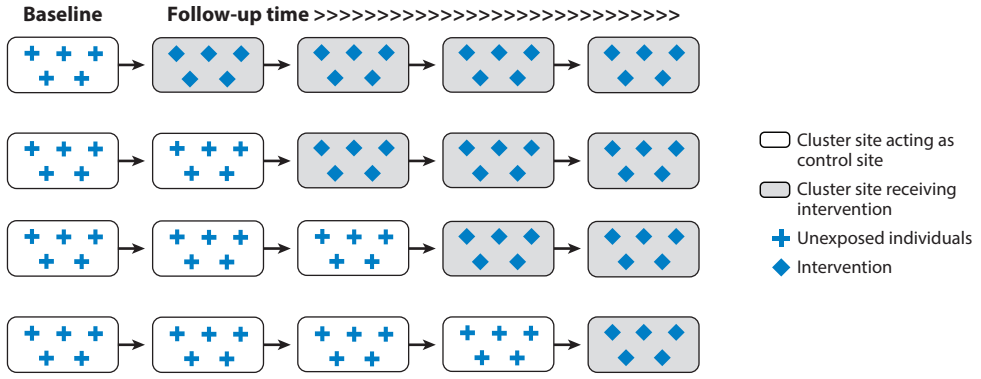

Figure 3

Illustration of the stepped-wedge study design. Intervention rollout over time.

Stepped-Wedge Designs

SWDs involve a sequential rollout of an intervention to participants (individuals or clusters) over several distinct time periods (5, 6, 21, 23, 28, 29, 38) (**Figure 3**). SWDs can include cohort designs (with the same individuals in each cluster in the pre- and postintervention steps) and repeated cross-sectional designs (with different individuals in each cluster in the pre- and post-intervention steps) (6). In the SWD, there is a unidirectional, sequential rollout of an intervention to clusters (or individuals) that occurs over different time periods. Initially, all clusters (or individuals) are unexposed to the intervention, and then, at regular intervals, selected clusters cross over (or step) into a time period where they receive the intervention. All clusters receive the intervention by the last time interval (although not all individuals within clusters necessarily receive the intervention). Data are collected on all clusters such that each cluster contributes data during both control and intervention time periods. The order in which clusters receive the intervention can be assigned randomly or using some other approach when randomization is not possible. For example, in settings with geographically remote or difficult-to-access populations, a nonrandom order can maximize efficiency with respect to logistical considerations.

The practical and social benefits of the SWD have been summarized in recent reviews (5, 21, 23, 26, 28, 36, 38, 42, 43, 47, 48, 53). In addition to addressing the general concerns with RCTs, as discussed above, the advantages of SWDs include the logistical convenience of staggering the intervention's rollout, which enables a smaller staff to be distributed across different implementation start times and allows for multilevel interventions to be integrated into practice or real-world settings (referred to as the feasibility benefit). This benefit also applies to studies of deimplementation, prior to a new approach being introduced. For example, with a staggered rollout, investigators can build in a transition cohort, such that sites can adjust to the integration of the new intervention or similarly allow for a staggered deimplementation of an existing practice. For a specified time period, there may be mixed or incomplete data, which can be designated as a washout period and excluded from the data analysis. However, extending rollout duration for practical reasons, such as this switching, may increase associated costs in threats to internal validity, as discussed below.

The SWD has several limitations. These generally involve consequences of the trade-offs related to having design control for the intervention rollout. This is often related to logistical reasons initially, but then can result in down-the-road threats to internal validity. These rollout-related threats include potential lagged intervention effects for nonacute outcomes; possible fatigue

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

and associated higher dropout rates among clusters assigned to receive the intervention later than others; fidelity losses for key intervention components over time; and potential contamination of later clusters (21). Another drawback of the SWD is that it involves data assessment at each point when a new cluster receives the intervention, substantially increasing the burden of data collection and costs unless data collection can be automated or uses existing data sources. Because the SWD often has more clusters receiving the intervention toward the end of the intervention period than in previous time periods, there is concern about potential temporal confounding at this stage. The SWD is also not as suited for evaluating intervention effects on delayed health outcomes (such as chronic disease incidence) and is most appropriate when outcomes occur relatively soon after each cluster starts receiving the intervention. Finally, as logistical necessity often dictates selecting a design with smaller numbers of clusters, there are related challenges in the statistical analysis. To use standard software, the common recommendation is to have at least 20–30 clusters (35).

SWDs can embed improvements that can enhance internal validity, mimicking the strength of RCTs. These generally focus on efforts to reduce bias or achieve balance in covariates across sites and over time and/or to compensate as much as possible for practical decisions made at the implementation stage, which affect the distribution of the intervention over time and by sites. The most widely used approaches are discussed in order of benefit to internal validity: (*a*) partial randomization; (*b*) stratification and matching; (*c*) embedding of data collection at critical points in time, such as with a phasing-in of intervention components, and (*d*) creation of a transition cohort or washout period. The most important of these SWD elements is random assignment of clusters into the intervention period. As well, utilizing data regarding time-varying covariates/confounders, either to stratify clusters and then randomize within strata (partial randomization) or to match clusters on known covariates in the absence of randomization, are techniques often employed to minimize bias and reduce confounding. Finally, maintaining control over the number and timing of data collection points over the study period can be beneficial in several ways. First, it can allow for data analysis strategies that can incorporate cyclical temporal trends (such as seasonality-mediated risk for the outcome, with flu or malaria) or other underlying temporal trends. Second, it can enable phased interventions to be studied for the contribution of different components included in the phases (e.g., passive then active intervention components) or can enable investigators to pause time, as when there is a structured washout or transition cohort created for practical reasons (e.g., one intervention or practice is stopped/deimplemented, and a new one is introduced) (see **Figure 4**).

Table 2 provides examples of studies using SWD that have used one or more of the design approaches described above to improve the internal validity of the study. In the study by Killam et al. (30), a nonrandomized SWD was used to evaluate a complex clinic-based intervention for integrating antiretroviral (ART) treatment into routine antenatal care in Zambia for postpartum

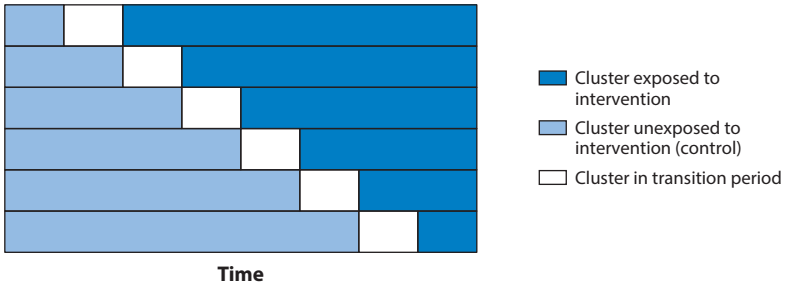

Figure 4

Illustration of the stepped-wedge study design. Summary of exposed and unexposed cluster time.

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

women. The design involved matching clinics by size and utilizing an inverse rollout to balance out the sizes across the four groups. The inverse roll-out involved four strata of clinics, grouped by size, with two clinics in each stratum. The rollout was sequenced across these eight clinics, such that one smaller clinic began early and three clinics of increasing size received the intervention afterward. This process was then followed by a descending order of clinics by size for the remaining rollout, ending with the smallest clinic. This inverse rollout enabled the investigators to start with a smaller clinic, to work out the logistical considerations, but then influence the rollout in order to avoid clustering of smaller or larger clinics in any one step of the intervention.

A second design feature of this study involved the use of a transition cohort or washout period (see **Figure 4**) (also used in the Morrison et al. study; 18, 37). This approach can be used when an existing practice is being replaced with the new intervention, but there is ambiguity about which group an individual would be assigned to while integration efforts were under way. In the Killam study, the concern was regarding women who might be identified as ART-eligible in the control period but then enroll in and initiate ART at an antenatal clinic during the intervention period. To account for the ambiguity of this transition period, researchers excluded patients with an initial antenatal visit more than 60 days prior to the date of implementing the ART in the intervention sites. For analysis of the primary outcome, patients were categorized into three mutually exclusive categories: a referral to the ART cohort, an integrated ART in the antenatal clinics cohort, and a transition cohort. It is important to note that incorporating time for a transition cohort can add considerable time to an intervention rollout, especially when the plan is also to deimplement an existing practice that involves a wide range of staff or activities. As well, the exclusion of the data during this phase can reduce the study's power if not built into the sample size considerations at the design phase.

Morrison et al. (37) used a randomized cluster design, with additional stratification and randomization within relevant subgroups, to examine a two-part quality improvement intervention focusing on clinician uptake of patient cooling procedures for postcardiac care in hospital settings (referred to as targeted temperature management). In this study, 32 hospitals were stratified into two groups on the basis of intensive care unit size ($<$10 beds versus $\geq$10 beds) and then randomly assigned into four different time periods to receive the intervention. The phased intervention implementation included both passive (generic didactic training components regarding the intervention) and active (tailored support to site-specific barriers identified in passive phase) components. This study exemplifies some of the best uses of SWD in the context of quality improvement interventions that have multiple components or for which there may be a passive and active phase, as is often the case with interventions that are layered onto systems change requirements (e.g., electronic records improvements/customization) or that relate to sequenced guidelines implementation (as in this example).

Studies using a wait-list partial randomization design are also included in **Table 2** (23, 26, 43). These types of studies are well suited to settings where there is routine enumeration of a cohort based on specific eligibility criteria, such as enrollment in a health plan or employment group, or inclusion in a disease-based registry (26, 43). This design can increase efficiency and statistical power in contrast with cluster-based trials, a crucial consideration when the number of participating individuals or groups is small (21).

The study by Grant et al. (23) uses a variant of the SWD for which individuals within a setting are enumerated and then randomized to get the intervention. In this example, employees who had previously screened positive for HIV at the company clinic as part of mandatory testing were invited in random sequence to attend a workplace HIV clinic at a large mining facility in South Africa to initiate a preventive medication treatment for tuberculosis (TB) (this occurred prior to the time when ARTs were more widely available). Individuals contributed follow-up time to the

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

preclinic phase from the baseline date established for the cohort until the actual date of their first clinic visit and also to the post-clinic phase thereafter. Clinic visits every six months were used to identify incident TB events. Because investigators were attempting to reduce TB incidence among the workers at the mine and not just those in the study, the effect of the intervention (the provision of clinic services) was estimated for the entire study population (incidence rate ratio).

CONSIDERATIONS IN CHOOSING AMONG QEDs

We present a decision map approach (**Figure 5**) to assist in the process of selecting among QEDs and demonstrate which features are emphasized in each design.

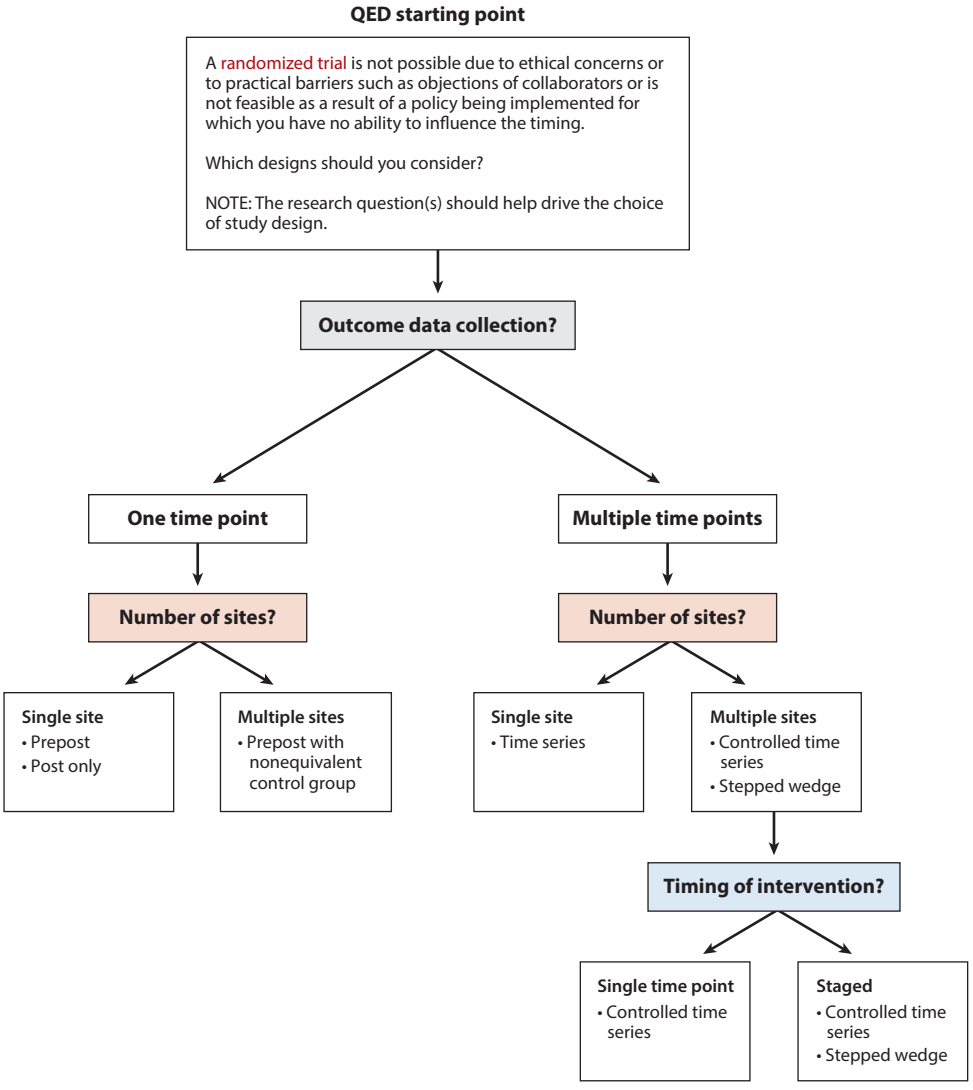

Figure 5

Quasi-experimental design (QED) decision-making map.

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

First, at the top of the flow diagram [1], one should consider that if there are multiple time points, then data can be collected for the pre- and postintervention periods. Ideally, more than two time points should be selected. If the inclusion of more than two time points is not possible, then multiple sites would allow for a nonequivalent prepost design. If more than two time points is possible for the study assessments, then one next should determine if multiple sites can be included [2]. If not, then one can consider a single-site-ITS. If multiple sites are possible, then one can choose between a SWD and a multiple-site ITS on the basis of whether the rollout will be observed over multiple time points (as in a SWD) or if only one intervention time point is possible (as in a controlled multiple-site ITS).

STRATEGIES TO STRENGTHEN EXTERNAL VALIDITY

In a recent article in the *Annual Review of Public Health* (24), the authors observed that there is an unavoidable trade-off between external and internal validity such that with higher control of a study, there is stronger evidence for internal validity but that control may jeopardize some of the external validity of the resultant stronger evidence. Nonetheless, several design strategies for nonexperimental studies can be utilized to improve internal validity without eliminating considerations of external validity. These are described below across all three study designs.

Examine Variation of Acceptability and Reach Among Diverse Subpopulations

One of the strengths of QEDs is that they are often employed to examine intervention effects in real-world settings and, often, for more diverse populations and settings. Consequently, if there is adequate examination of participants' characteristics and setting-related factors, it can be possible to interpret findings among critical groups for which there may be no existing evidence of an intervention effect. For example, in the Campus Watch intervention (14), the investigator oversampled the Maori indigenous population to be able to stratify the results and investigate whether the program was effective for this understudied group. In the study by Zombré et al. (54) on health care access in Burkina Faso, the authors examined clinic density characteristics to determine their impact on sustainability.

Characterize Fidelity and Measures of Implementation Processes

Some of the most important outcomes for examination in these QED studies include whether the intervention was delivered as intended (i.e., fidelity), whether it was maintained over the entire study period (i.e., sustainability), and if the outcomes could be examined specifically by this level of fidelity within or across sites. As well, when a complex intervention is related to a policy or guideline shift and implementation requires logistical adjustments (such as phased rollouts to embed the intervention or to train staff), QEDs more truly mimic real-world constraints. As a result, capturing processes of implementation are critical because they can describe important variation in uptake, informing interpretation of the findings for external validity. As described by Prost et al. (42), for example, it is essential to capture what occurs during such phased intervention rollouts, as with following established guidelines for the development of complex interventions, including efforts to define and protocolize activities before their implementation (15, 16, 27). However, QEDs are often conducted by teams with strong interests in adapting the intervention or learning by doing, which can limit interpretation of findings if not planned into the design. As done in the study by Bailet et al. (3), the investigators refined the intervention, based on year 1 data, and then applied the modified intervention in years 2 and 3, collecting additional data on training

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

and measurement fidelity in later years. This phasing aspect of implementation generates tension between protocolizing interventions and adapting them as they go along. When adaptation is necessary, additional designs for the intervention rollout, such as adaptive or hybrid designs, can also be considered.

Conduct Community or Cohort-Based Sampling to Improve Inference

External validity can be improved when the intervention is applied to entire communities, as with some of the community-randomized studies described in **Table 2** (11, 20). In these cases, the results are closer to the conditions that would apply if the interventions were conducted at scale, with a large proportion of a population receiving the intervention. In some cases, QEDs also afford greater access for some intervention research to be conducted in remote or difficult-to-reach communities, where the cost and logistical requirements of an RCT may become prohibitive or may require alteration of the intervention or staffing support to levels that would never be feasible in real-world application.

Employ a Model or Framework that Covers Both Internal and External Validity

Frameworks can be helpful to enhance interpretability of many kinds of studies, including QEDs, and can help ensure that information on essential implementation strategies is included in the results (46). Although several of the case studies summarized in this article included measures that can improve external validity (such as subgroup analysis of which participants were most impacted as well as processes and contextual measures that can affect variation in uptake), none formally employ an implementation framework. Green & Glasgow (25) have outlined several useful criteria for gauging the extent to which an evaluation study also provides measures that enhance interpretation of external validity; researchers and practitioners employing QEDs could use such a framework to identify relevant components to include in reported findings.

CONCLUSION

It is more difficult to conduct a good quasi-experiment than to conduct a good randomized trial (45). Although QEDs are increasingly used, it is important to note that randomized designs are still preferred over quasi-experiments except where randomization is not possible. In this review, we present three important QEDs and variants nested within them that can increase internal validity while also improving external validity considerations, and we present case studies employing these techniques.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This publication was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through UCSF-CTSI grant number UL1 TR001872. The article's contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

LITERATURE CITED

1. Alonso PL, Lindsay SW, Armstrong-Schellenberg JRM, Konteh M, Keita MK, et al. 1993. Special Issue: Malaria Control Trial Using Insecticide-Treated Bed Nets and Targeted Chemoprophylaxis in a Rural Area of The Gambia, West Africa. *Trans. R. Soc. Trop. Med. Hyg.* 87(2):1–60
2. Ammerman A, Smith T, Calancie L. 2014. Practice-based evidence in public health: improving reach, relevance, and results. *Annu. Rev. Public Health* 35:47–63
3. Bailet LL, Repper K, Murphy S, Piasta S, Zettler-Greeley C. 2013. Emergent literacy intervention for pre-kindergarteners at risk for reading failure: years 2 and 3 of a multiyear study. *J. Learn. Disabil.* 46:133–53
4. Basu S, Meghani A, Siddiqi A. 2017. Evaluating the health impact of large-scale public policy changes: classical and novel approaches. *Annu. Rev. Public Health* 38:351–70
5. Beard E, Lewis JJ, Copas A, Davey C, Osrin D, et al. 2015. Stepped wedge randomised controlled trials: systematic review of studies published between 2010 and 2014. *Trials* 16:353
6. Brown CA, Lilford RJ. 2006. The stepped wedge trial design: a systematic review. *BMC Med. Res. Methodol.* 6:54
7. Brown CH, Curran G, Palinkas LA, Aarons GA, Wells KB, et al. 2017. An overview of research and evaluation designs for dissemination and implementation. *Annu. Rev. Public Health* 38:1–22
8. Brown CH, Ten Have TR, Jo B, Dagne G, Wyman PA, et al. 2009. Adaptive designs for randomized trials in public health. *Annu. Rev. Public Health* 30:1–25
9. Brownson RC, Diez Roux AV, Swartz K. 2014. Commentary: Generating rigorous evidence for public health: the need for new thinking to improve research and practice. *Annu. Rev. Public Health* 35:1–7
10. Campbell DT, Stanley JC. 1963. Experimental and quasi-experimental designs for research on teaching. In *Handbook of Research on Teaching*, ed. NL Gage. Chicago: Rand McNally
11. Cissé B, Ba EH, Sokhna C, NDiaye JL, Gomis JF, et al. 2016. Effectiveness of seasonal malaria chemoprevention in children under ten years of age in Senegal: a stepped-wedge cluster-randomised trial. *PLOS Med.* 13:e1002175
12. Cochrane Effective Practice Organ. Care (EPOC). 2017. *Interrupted time series (ITS) analyses. EPOC resources for review authors.* EPOC, London. **http://epoc.cochrane.org/epoc-specific-resources-review-authors**
13. Cook TD, Campbell DT. 1976. The design and conduct of true experiments and quasi-experiments in field settings. In *Handbook of Industrial and Organizational Psychology*, ed. MD Dunnette, pp. 223–326. Washington, DC: Am. Psychol. Assoc.
14. Cousins K, Connor JL, Kypri K. 2014. Effects of the Campus Watch intervention on alcohol consumption and related harm in a university population. *Drug Alcohol Depend.* 143:120–26
15. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. 2008. Developing and evaluating complex interventions: the new Medical Research Council guidance. *Br. Med. J.* 337:1655
16. Craig P, Katikireddi SV, Leyland A, Popham F. 2017. Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. *Annu. Rev. Public Health* 38:39–56
17. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. 2012. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Med. Care* 50(3):217–26
18. Dainty KN, Scales D, Brooks SC, Needham DM, Dorian P, et al. 2011. A knowledge translation collaborative to improve the use of therapeutic hypothermia in post-cardiac arrest patients: protocol for a stepped wedge randomized trial. *Implementation Sci.* 6:4
19. Fan E, Laupacis A, Pronovost PJ, Guyatt GH, Needham DM. 2010. How to use an article about quality improvement. *JAMA* 304:2279–87
20. Fernald LCH, Gertler PJ, Neufeld LM. 2008. Role of cash in conditional cash transfer programmes for child health, growth, and development: an analysis of Mexico's *Oportunidades*. *Lancet* 371:828–37
21. Fok CCT, Henry D, Allen A. 2015. Research designs for intervention research with small samples II: stepped wedge and interrupted time-series designs. *Prev. Sci.* 16:967–77
22. Glasgow RE, Vinson C, Chambers D, Khoury MJ, Kaplan RM, Hunter C. 2012. National Institutes of Health approaches to dissemination and implementation science: current and future directions. *Am. J. Public Health* 102(7):1274–81

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

23. Grant AD, Charalambous S, Fielding KL, Day JH, Corbett EL, et al. 2005. Effect of routine isoniazid preventive therapy on tuberculosis incidence among HIV-infected men in South Africa: a novel randomized recruitment study. *JAMA* 293:2719–25
24. Green LW, Brownson RC, Fielding JE. 2017. Introduction: How is the growing concern for relevance and implementation of evidence-based interventions shaping the public health research agenda? *Annu. Rev. Public Health* 38:v–vii
25. Green LW, Glasgow RE. 2006. Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology. *Eval. Health Prof.* 29(1):126–53
26. Handley MA, Schillinger D, Shiboski S. 2011. Quasi-experimental designs in practice-based research settings: design and implementation considerations. *J. Am. Board Fam. Med.* 24:589–96
27. Hawe P. 2015. Lessons from complex interventions to improve health. *Annu. Rev. Public Health* 36:307–23
28. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. 2015. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *Br. Med. J.* 350:h391
29. Hussey MA, Hughes JP. 2007. Design and analysis of stepped wedge cluster randomized trials. *Contemp. Clin. Trials* 28:182–91
30. Killam WP, Tambatamba BC, Chintu N, Rouse D, Stringer E, et al. 2010. Antiretroviral therapy in antenatal care to increase treatment initiation in HIV-infected pregnant women: a stepped-wedge evaluation. *AIDS* 24:85–89
31. Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D. 2015. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *BMJ* 350:h2750
32. Krass I. 2016. Quasi experimental designs in pharmacist intervention research. *Int. J. Clin. Pharm.* 38:647–54
33. Leviton LC. 2017. Generalizing about public health interventions: a mixed-methods approach to external validity. *Annu. Rev. Public Health* 38:371–91
34. Lopez Bernal J, Cummins S, Gasparrini A. 2017. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int. J. Epidemiol.* 46(1):348–55
35. McNeish DM, Harring JR. 2014. Clustered data with small sample sizes: comparing the performance of model-based and design-based approaches. *Commun. Stat.* 46:855–69
36. Mdege ND, Man MS, Taylor CA, Torgerson DJ. 2011. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. *J. Clin. Epidemiol.* 64:936–48
37. Morrison LJ, Brooks SC, Dainty KN, Dorian P, Needham DM, et al. 2015. Improving use of targeted temperature management after out-of-hospital cardiac arrest: a stepped wedge cluster randomized controlled trial. *Crit. Care Med.* 43:954–64
38. Murray DM, Pennell M, Rhoda D, Hade EM, Paskett ED. 2010. Designing studies that would address the multilayered nature of health care. *J. Natl. Cancer Inst. Monogr.* 40:90–96
39. Naci H, Soumerai SB. 2016. History bias, study design, and the unfulfilled promise of pay-for-performance policies in health care. *Prev. Chronic. Dis.* 13:160133
40. Pellegrin KL, Krenk L, Oakes SJ, Ciarleglio A, Lynn J, et al. 2017. Reductions in medication-related hospitalizations in older adults by hospital and community pharmacists: a quasi-experimental study. *J. Am. Geriatr. Soc.* 65:212–19
41. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. 2013. Implementation research: what it is and how to do it. *BMJ* 347:f6753
42. Prost A, Binik A, Abubakar I, Roy A, De Allegri M, et al. 2015. Logistic, ethical, and political dimensions of stepped wedge trials: critical review and case studies. *Trials* 16:351
43. Ratanawongsa N, Handley MA, Quan J, Sarkar U, Pfeifer K, et al. 2012. Quasi-experimental trial of diabetes Self-Management Automated and Real-Time Telephonic Support (SMARTSteps) in a Medicaid managed care plan: study protocol. *BMC Health Serv. Res.* 12:22
44. Robinson TE, Zhou L, Kerse N, Scott JDR, Christiansen JP, et al. 2015. Evaluation of a New Zealand program to improve transition of care for older high risk adults. *Australas. J. Ageing* 34:269–74
45. Shadish WR, Cook TD, Campbell DT. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York: Houghton Mifflin

46. Tabak RG, Khoong EC, Chambers DA, Brownson RC. 2012. Bridging research and practice: models for dissemination and implementation research. *Am. J. Prev. Med.* 43(3):337–50
47. Turner EL, Li F, Gallis JA, Prague M, Murray DM. 2017. Review of recent methodological developments in group-randomized trials: part 1—design. *Am. J. Public Health* 107:907–15
48. Turner EL, Prague M, Gallis JA, Li F, Murray DM. 2017. Review of recent methodological developments in group-randomized trials: part 2—analysis. *Am. J. Public Health* 107:1078–86
49. Weinhardt LS, Galvao LW, Yan AF, Stevens P, Mwenyekonde TN, et al. 2017. Mixed-method quasi-experimental study of outcomes of a large-scale multilevel economic and food security intervention on HIV vulnerability in rural Malawi. *AIDS Behav*. 21:712–23
50. Wertz D, Hou L, Devries A, Dupclay L Jr., McGowan F, et al. 2012. Clinical and economic outcomes of the Cincinnati Pharmacy Coaching Program for diabetes and hypertension. *Manag. Care* 21:44–54
51. West SG, Duan N, Pequegnat W, Gaist P, Des Jarlais DC, et al. 2008. Alternatives to the randomized controlled trial. *Am. J. Public Health* 98:1359–66
52. White H, Sabarwal S. 2014. *Quasi-experimental design and methods*. Methodol. Briefs: Impact Eval. 8. UNICEF Off. Res., Florence. **https://www.unicef-irc.org/publications/pdf/brief_8_quasi-experimental%20design_eng.pdf**
53. Wyman PA, Henry D, Knoblauch S, Brown CH. 2015. Designs for testing group-based interventions with limited numbers of social units: the dynamic wait-listed and regression point displacement designs. *Prev. Sci.* 16:956–66
54. Zombré D, Allegri MD, Ridde V. 2017. Immediate and sustained effects of user fee exemption on health-care utilization among children under five in Burkina Faso: a controlled interrupted time-series analysis. *Soc. Sci. Med.* 179:27–35

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

ANNUAL REVIEWS
Connect With Our Experts

New From Annual Reviews:

Annual Review of Criminology

criminol.annualreviews.org • Volume 1 • January 2018

ONLINE NOW!

Co-Editors: **Joan Petersilia**, *Stanford University* and **Robert J. Sampson**, *Harvard University*

The *Annual Review of Criminology* provides comprehensive reviews of significant developments in the multidisciplinary field of criminology, defined as the study of both the nature of criminal behavior and societal reactions to crime. International in scope, the journal examines variations in crime and punishment across time (e.g., why crime increases or decreases) and among individuals, communities, and societies (e.g., why certain individuals, groups, or nations are more likely than others to have high crime or victimization rates). The societal effects of crime and crime control, and why certain individuals or groups are more likely to be arrested, convicted, and sentenced to prison, will also be covered via topics relating to criminal justice agencies (e.g., police, courts, and corrections) and criminal law.

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

TABLE OF CONTENTS FOR VOLUME 1:

THE DISCIPLINE

- *Reflections on Disciplines and Fields, Problems, Policies, and Life,* James F. Short
- *Replication in Criminology and the Social Sciences,* William Alex Pridemore, Matthew C. Makel, Jonathan A. Plucker

CRIME AND VIOLENCE

- *Bringing Crime Trends Back into Criminology: A Critical Assessment of the Literature and a Blueprint for Future Inquiry,* Eric P. Baumer, María B. Vélez, Richard Rosenfeld
- *Immigration and Crime: Assessing a Contentious Issue,* Graham C. Ousey, Charis E. Kubrin
- *The Long Reach of Violence: A Broader Perspective on Data, Theory, and Evidence on the Prevalence and Consequences of Exposure to Violence,* Patrick Sharkey
- *Victimization Trends and Correlates: Macro- and Microinfluences and New Directions for Research,* Janet L. Lauritsen, Maribeth L. Rezey
- *Situational Opportunity Theories of Crime,* Pamela Wilcox, Francis T. Cullen
- *Schools and Crime,* Paul J. Hirschfield

PUNISHMENT AND POLICY

- *Collateral Consequences of Punishment: A Critical Review and Path Forward,* David S. Kirk, Sara Wakefield
- *Understanding the Determinants of Penal Policy: Crime, Culture, and Comparative Political Economy,* Nicola Lacey, David Soskice, David Hope
- *Varieties of Mass Incarceration: What We Learn from State Histories,* Michael C. Campbell
- *The Politics, Promise, and Peril of Criminal Justice Reform in the Context of Mass Incarceration,* Katherine Beckett

THE PRISON

- *Inmate Society in the Era of Mass Incarceration,* Derek A. Kreager, Candace Kruttschnitt
- *Restricting the Use of Solitary Confinement,* Craig Haney

DEVELOPMENTAL AND LIFE-COURSE CRIMINOLOGY

- *Desistance from Offending in the Twenty-First Century,* Bianca E. Bersani, Elaine Eggleston Doherty
- *On the Measurement and Identification of Turning Points in Criminology,* Holly Nguyen, Thomas A. Loughran

ECONOMICS OF CRIME

- *Gun Markets,* Philip J. Cook
- *Offender Decision-Making in Criminology: Contributions from Behavioral Economics,* Greg Pogarsky, Sean Patrick Roche, Justin T. Pickett

POLICE AND COURTS

- *Policing in the Era of Big Data,* Greg Ridgeway
- *Reducing Fatal Police Shootings as System Crashes: Research, Theory, and Practice,* Lawrence W. Sherman
- *The Problems With Prosecutors,* David Alan Sklansky
- *Monetary Sanctions: Legal Financial Obligations in US Systems of Justice,* Karin D. Martin, Bryan L. Sykes, Sarah Shannon, Frank Edwards, Alexes Harris
- *Forensic DNA Typing,* Erin Murphy

ANNUAL REVIEWS | CONNECT WITH OUR EXPERTS
650.493.4400/800.523.8635 (US/CAN)
www.annualreviews.org | service@annualreviews.org

Annual Review of Public Health
Volume 39, 2018

Contents

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

Symposium

Commentary: Increasing the Connectivity Between Implementation Science and Public Health: Advancing Methodology, Evidence Integration, and Sustainability
David A. Chambers 1

Selecting and Improving Quasi-Experimental Designs in Effectiveness and Implementation Research
Margaret A. Handley, Courtney R. Lyles, Charles McCulloch, and Adithya Cattamanchi 5

Building Capacity for Evidence-Based Public Health: Reconciling the Pulls of Practice and the Push of Research
Ross C. Brownson, Jonathan E. Fielding, and Lawrence W. Green 27

The Sustainability of Evidence-Based Interventions and Practices in Public Health and Health Care
Rachel C. Shelton, Brittany Rhoades Cooper, and Shannon Wiltsey Stirman 55

Epidemiology and Biostatistics

Selecting and Improving Quasi-Experimental Designs in Effectiveness and Implementation Research
Margaret A. Handley, Courtney R. Lyles, Charles McCulloch, and Adithya Cattamanchi 5

Agent-Based Modeling in Public Health: Current Applications and Future Directions
Melissa Tracy, Magdalena Cerdá, and Katherine M. Keyes 77

Big Data in Public Health: Terminology, Machine Learning, and Privacy
Stephen J. Mooney and Vikas Pejaver 95

Environmental Determinants of Breast Cancer
Robert A. Hiatt and Julia Green Brody 113

Meta-Analysis of Complex Interventions
Emily E. Tanner-Smith and Sean Grant 135

Precision Medicine from a Public Health Perspective
Ramya Ramaswami, Ronald Bayer, and Sandro Galea 153

Relative Roles of Race Versus Socioeconomic Position in Studies of Health Inequalities: A Matter of Interpretation
Amani M. Nuru-Jeter, Elizabeth K. Michaels, Marilyn D. Thomas, Alexis N. Reeves, Roland J. Thorpe Jr., and Thomas A. LaVeist 169

Social Environment and Behavior

The Debate About Electronic Cigarettes: Harm Minimization or the Precautionary Principle
Lawrence W. Green, Jonathan E. Fielding, and Ross C. Brownson 189

Harm Minimization and Tobacco Control: Reframing Societal Views of Nicotine Use to Rapidly Save Lives
David B. Abrams, Allison M. Glasser, Jennifer L. Pearson, Andrea C. Villanti, Lauren K. Collins, and Raymond S. Niaura 193

E-Cigarettes: Use, Effects on Smoking, Risks, and Policy Implications
Stanton A. Glantz and David W. Bareham 215

Increasing Disparities in Mortality by Socioeconomic Status
Barry Bosworth 237

Neighborhood Interventions to Reduce Violence
Michelle C. Kondo, Elena Andreyeva, Eugenia C. South, John M. MacDonald, and Charles C. Branas 253

The Relationship Between Education and Health: Reducing Disparities Through a Contextual Approach
Anna Zajacova and Elizabeth M. Lawrence 273

Environmental and Occupational Health

Building Evidence for Health: Green Buildings, Current Science, and Future Challenges
J.G. Cedeño-Laurent, A. Williams, P. MacNaughton, X. Cao, E. Eitland, J. Spengler, and J. Allen 291

Environmental Influences on the Epigenome: Exposure-Associated DNA Methylation in Human Populations
Elizabeth M. Martin and Rebecca C. Fry 309

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

From Crowdsourcing to Extreme Citizen Science: Participatory Research for Environmental Health
P.B. English, M.J. Richardson, and C. Garzón-Galvis 335

Migrant Workers and Their Occupational Health and Safety
Sally C. Moyce and Marc Schenker 351

Mobile Sensing in Environmental Health and Neighborhood Research
Basile Chaix 367

Public Health Practice and Policy

Commentary: Increasing the Connectivity Between Implementation Science and Public Health: Advancing Methodology, Evidence Integration, and Sustainability
David A. Chambers 1

Building Capacity for Evidence-Based Public Health: Reconciling the Pulls of Practice and the Push of Research
Ross C. Brownson, Jonathan E. Fielding, and Lawrence W. Green 27

The Sustainability of Evidence-Based Interventions and Practices in Public Health and Health Care
Rachel C. Shelton, Brittany Rhoades Cooper, and Shannon Wiltsey Stirman 55

The Debate About Electronic Cigarettes: Harm Minimization or the Precautionary Principle
Lawrence W. Green, Jonathan E. Fielding, and Ross C. Brownson 189

Harm Minimization and Tobacco Control: Reframing Societal Views of Nicotine Use to Rapidly Save Lives
David B. Abrams, Allison M. Glasser, Jennifer L. Pearson, Andrea C. Villanti, Lauren K. Collins, and Raymond S. Niaura 193

E-Cigarettes: Use, Effects on Smoking, Risks, and Policy Implications
Stanton A. Glantz and David W. Bareham 215

Neighborhood Interventions to Reduce Violence
Michelle C. Kondo, Elena Andreyeva, Eugenia C. South, John M. MacDonald, and Charles C. Branas 253

Mobile Sensing in Environmental Health and Neighborhood Research
Basile Chaix 367

Policy Approaches for Regulating Alcohol Marketing in a Global Context: A Public Health Perspective
Marissa B. Esser and David H. Jernigan 385

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.

Problems and Prospects: Public Health Regulation of Dietary Supplements
Colin W. Binns, Mi Kyung Lee, and Andy H. Lee 403

Health Services

Achieving Mental Health and Substance Use Disorder Treatment Parity: A Quarter Century of Policy Making and Research
Emma Peterson and Susan Busch 421

Data Resources for Conducting Health Services and Policy Research
Lynn A. Blewett, Kathleen Thiede Call, Joanna Turner, and Robert Hest 437

Designing Difference in Difference Studies: Best Practices for Public Health Policy Research
Coady Wing, Kosali Simon, and Ricardo A. Bello-Gomez 453

How Much Do We Spend? Creating Historical Estimates of Public Health Expenditures in the United States at the Federal, State, and Local Levels
Jonathon P. Leider, Beth Resnick, David Bishai, and F. Douglas Scutchfield 471

Modeling Health Care Expenditures and Use
Partha Deb and Edward C. Norton 489

Promoting Prevention Under the Affordable Care Act
Nadia Chait and Sherry Glied 507

Treatment and Prevention of Opioid Use Disorder: Challenges and Opportunities
Dennis McCarty, Kelsey C. Priest, and P. Todd Korthuis 525

Indexes

Cumulative Index of Contributing Authors, Volumes 30–39 543

Cumulative Index of Article Titles, Volumes 30–39 549

Errata

An online log of corrections to *Annual Review of Public Health* articles may be found at http://www.annualreviews.org/errata/publhealth

Annu. Rev. Public Health 2018.39:5-25. Downloaded from www.annualreviews.org
Access provided by 206.57.72.13 on 06/13/18. See copyright for approved use.