Evaluation of Systems-Oriented Public Health Interventions: Alternative Research Designs

Robert W. Sanson-Fisher,[1] Catherine A. D'Este,[2] Mariko L. Carey,[1] Natasha Noble,[1] and Christine L. Paul[1]

[1]Priority Research Center for Health Behavior and Hunter Medical Research Institute, University of Newcastle, Callaghan, NSW 2308, Australia; email: Rob.Sanson-Fisher@newcastle.edu.au, Mariko.Carey@newcastle.edu.au, Natasha.Noble@newcastle.edu.au, Chris.Paul@newcastle.edu.au

[2]School of Medicine and Public Health, University of Newcastle, Callaghan, NSW 2308, Australia; email: Catherine.dEste@newcastle.edu.au

Keywords

cluster randomized controlled trial, stepped wedge design, interrupted time series, multiple baseline design, controlled prepost design

Abstract

The need to provide sound evidence of the costs and benefits of real-world public health interventions has driven advances in the development and analysis of designs other than the controlled trial in which individuals are randomized to an experimental condition. Attention to methodological quality is of critical importance to ensure that any evaluation can accurately answer three fundamental questions: (*a*) Has a change occurred, (*b*) did the change occur as a result of the intervention, and (*c*) is the degree of change significant? A range of alternatives to the individual randomized controlled trial (RCT) can be used for evaluating such interventions, including the cluster RCT, stepped wedge design, interrupted time series, multiple baseline, and controlled prepost designs. The key features and complexities associated with each of these designs are explored.

INTRODUCTION

Public health interventions are strategies designed to improve the health of the population as a whole, or particular subgroups within a population (5, 79). Interventions may come in a variety of formats ranging from government-funded vaccinations for common infectious diseases (52, 89, 95) to programs to improve nutrition among school children (54, 68) to smoking cessation campaigns (27, 94).

Evaluation of Public Health Interventions Is Critical to Ensuring that Interventions Are Effective and Not Harmful

Public health interventions must be rigorously evaluated for several reasons. In addition to ensuring that public health funds are spent in the most effective manner possible, researchers need to ensure that interventions that are potentially ineffective or harmful are not implemented.

Unfortunately, we note many examples where public health interventions have been implemented without a strong evidence base. For example, in 1997 the World Health Organization (WHO) recommended the directly observed treatment, short course (DOTS) strategy, which used independent and trained personnel to observe tuberculosis patients taking their medication as a means to improve treatment adherence (98). A subsequent Cochrane systematic review conducted in 2007 found no evidence that direct observation resulted in greater tuberculosis cure rates than self-administered treatment (92). Given the expense involved in direct observation, the recommendation of this method despite a lack of evidence is concerning. Similarly, many men undergo prostate cancer screening [33% of men over the age of 40 (34) and 41% of men aged 50 and older (88)] in the United States despite insufficient evidence to support population-based screening for prostate cancer (23, 100). Negative potential effects of screening include a high false-positive rate (up to 76%), risk of infection, bleeding and pain with biopsy (49), and overtreatment (99). Such examples highlight the importance of ensuring a strong evidence base for public health interventions.

Evaluation Is Needed to Determine Effects of Interventions on Health Inequalities

In addition to ensuring that public health interventions are effective, we need to assess the impact on health equity. Substantial inequalities in health outcomes are associated with factors such as socioeconomic status (59) and cultural heritage (51). For example, Marmot (2005) reports a spread of life expectancy of 48 years between countries, and 20 years within countries; Aboriginal and Torres Strait Islanders experience significantly poorer life expectancy than does the general Australian population (60). Similarly, disparities in life expectancy across US counties have been increasing, largely attributable to a plateau or decline in life expectancy among the most disadvantaged segments of the population (28). Such disparities in life expectancy provide an indication that certain subgroups of a population are less likely to receive access to health interventions or are less likely to fully benefit when access is provided.

In the case of smoking, analysis of national health survey data from 18 European countries, all with some form of national tobacco control policy, revealed a social gradient such that smoking prevalence is higher among those with lower levels of education (83). Smokers with a higher education and higher income are more likely to intend to quit, to make a quit attempt, and to be abstinent for at least one month relative to smokers with lower education and income (76). These discrepancies in health behaviors and outcomes reinforce the need for careful, rigorous, and thorough evaluation of public health efforts, including their impact on both overall population

health and vulnerable population groups (72). Thus we need to understand not only the efficacy of interventions under tightly controlled conditions, but also the programs' effectiveness across settings and populations.

This review provides an overview of the principles of rigorous evaluation of public health interventions using the randomized controlled trial (RCT) design; describes the circumstances driving the need to consider alternative designs to the RCT; and describes selected alternative designs to the RCT, along with examples of how these designs have been used to evaluate public health interventions.

PRINCIPLES FOR EVALUATING PUBLIC HEALTH INITIATIVES

Given the need to evaluate public health initiatives to maximize health benefits, minimize harms, avoid exacerbation of health inequities, and maximize the value of resources, the question of what constitutes evidence of effectiveness is critical. In public health, approaches to evaluation include process evaluation (82), action research (73), and realist evaluation (71). Although they are valuable for the assessment of complex interventions, these approaches generally have shortcomings in the extent to which an effect can be attributed to the intervention, whether findings may be generalized to other settings, and whether intervention effects are quantified (24, 71).

Frameworks such as those developed by the Cochrane Collaboration (45), US Preventive Services Task Force (39), US Task Force on Community Preventive Services (101), and Center for Evidence-Based Medicine (17), among others, are internationally accepted approaches for assessing the strength of evidence for health interventions. These frameworks were developed to organize medical and public health research information in a systematic way. In these frameworks, the RCT design is considered the gold standard for evaluating intervention effectiveness. The RCT is grounded in a medical paradigm in which individuals are randomly assigned to receive an intervention or a control condition. Randomization is central because it eliminates selection bias, i.e., where intervention assignment might be influenced by perceived relevance or likely benefit. Randomization also generally balances groups on variables (both known and unknown) that might affect the outcome of the study; thus any differences in outcomes observed between groups can be attributed to the intervention rather than to another cause (87). Such frameworks also recognize that one intervention study does not necessarily provide proof of effectiveness and that the strongest evidence is derived from consistent findings across studies (32). This view has led to the establishment of methodologies for synthesizing evidence across studies in systematic reviews (45). However, although the RCT, with random assignment of individuals, is well suited to the evaluation of discrete individualized medical treatments, this method has limitations when applied to public health intervention evaluation, where it is often necessary to find an appropriate balance between study design and feasibility or external validity (63, 80).

AN INCREASING RECOGNITION IN PUBLIC HEALTH THAT CHANGING BEHAVIORS REQUIRES SYSTEMS INTERVENTIONS

The focus of public health efforts is increasingly moving from infectious diseases to prevention of chronic disease (96). Chronic diseases arise in part through a range of modifiable behaviors such as tobacco smoking, physical inactivity, unhealthy diet, and harmful use of alcohol (97). These complex behaviors are in turn influenced by a range of individual, social, and environmental factors (70). This focus on complex behaviors has posed a challenge for public health researchers to develop interventions that are multifactorial and can address the many influences on behavior. In particular, rather than changing individual attitudes, knowledge, or behavior, interventions may need to change aspects of whole systems. Systems relevant to public health interventions

include physician practices, hospitals, geographically defined communities or regions, social groups, and cultural groups.

AN UNDERSTANDING OF THE COMPLEX RANGE OF FACTORS THAT INFLUENCE BEHAVIOR

Alongside the change in focus regarding the types of behavior targeted by public health interventions, a more sophisticated understanding of human behavior and its antecedents has also developed. Some decades ago this area was dominated by theorists such as Freud (29), Rogers (77), and Adler (1), who maintained that behavior could be changed through intrapsychic examination and therapy. Later models of behavior continued to have an individual focus in explaining health behavior such as social learning theory (6), the health belief model (50), and the theory of planned behavior (2).

Current frameworks for conceptualizing and changing human behavior—particularly those on the large scale needed to achieve population-level benefit—incorporate a layered perspective, recognizing the influence of health systems and social factors on public health. For example, the PRECEDE-PROCEED model (33) and the behavior change wheel (64) address factors operating at the individual, social, and system levels. Models of the socioeconomic determinants of health in particular (60) place a strong emphasis on factors external to the individual in driving health outcomes. Relatively mature fields of public health research (e.g., tobacco control) are associated with the use of a range of multifaceted approaches in public health endeavors, including taxation, community-level regulation, mass media campaigns, and individual cessation support strategies (e.g., 21). As public health interventions are moving toward whole-systems or multilayered interventions, an evaluation approach that can accommodate and assess the effect of such interventions is needed.

EVALUATING CHANGE IN COMPLEX SYSTEMS CAN CONFLICT WITH RANDOMIZATION OF INDIVIDUALS

Designs that randomly assign individuals to intervention or control conditions are generally incompatible with systems-based interventions; RCTs often cannot accommodate the complexity and flexibility required to evaluate interventions at the systems or policy level (e.g., 7, 63, 102). However, researchers may be reluctant to use alternatives to the RCT, owing to concerns that the alternatives will not be accepted by funding agencies, peer reviewers, or the groups or systems of interest. Researcher training and experience are also likely to influence the choice of alternative research designs. The tension between design rigor and potentially effective systems interventions can result in poor or nonexistent evaluation of public health initiatives, or alternatively in failure to implement potentially effective systems-based initiatives. Rather than selecting interventions that can be evaluated using an RCT, we should be selecting those interventions that are most likely to produce the desired change and should consider which evaluation design is appropriate for determining whether this intervention is effective. Given the limitations of standard, randomized approaches, we must explore the degree to which alternative designs to the RCT can be used to deliver rigorous evaluation methodology for real-world, whole-systems public health initiatives.

THE RESEARCH DESIGN SHOULD ALLOW THREE BASIC QUESTIONS TO BE ANSWERED

Whether an intervention is simple or complex, methodological quality-assessment tools, such as the Cochrane Risk of Bias assessment (44), reinforce the fundamental importance of error

elimination such as those due to selection bias and measurement bias. Methodological quality is critically important to ensure that the research design employed to evaluate a treatment or intervention can accurately answer three fundamental questions: (*a*) Has a change occurred, (*b*) did the change occur as a result of the intervention and not some extraneous factor or cause, and (*c*) is the degree of change perceived to be significant to important stakeholders?

Has a Change Occurred?

The primary goal of public health interventions is to improve important health outcomes; therefore, evaluation of interventions must be able to establish whether a desired change in health has occurred. Often decision makers assume that positive change will occur if resources are invested in public health. One example is the detection of patients' risk-factor status by primary care physicians. Targeted action over the past decade has aimed at changing the complex factors believed to influence risk-factor detection rates and, therefore, the adequacy of preventive care offered. Investment in improving detection has included changes in financial compensation for primary care physicians who engage in preventive activities (78), the provision of nurse practitioners in surgeries to assist with aspects of clinical care (47), and the promotion of computerized medical record systems to encourage routine screening and recording of risk factors (46).

Despite such targeted and well-resourced actions, rates of risk-factor detection in primary care patients have not changed significantly over the past three decades: Several studies exploring general practitioners' detection of smoking and high-risk drinking among their Australian patients conducted from the 1980s to 2012 found little difference in general practitioner detection of health risks over this time period. Studies reported that primary care physicians fail to detect ~40% of patients who were smokers (22, 43; S. Yoong, unpublished data) and 70% of individuals who consumed alcohol at risky levels (75; C. Paul, under editorial review). Substantive change has not occurred in detection rates of these health-risk behaviors by primary care physicians in spite of considerable financial and human investment.

In contrast, public health interventions such as those using mass media have been credited with positive change in relation to tobacco, alcohol, cardiovascular risk, and cancer screening (93). In the United States, adult smoking prevalence has declined from 33% in 1980 to 19% in 2010 (15, 16). Similar change has occurred in Australia, where smoking prevalence has declined from ~40% of adults in 1977 to 15% in 2010 (3, 4). While these findings clearly indicate that changes have occurred in the prevalence of smoking, demonstrating that change has occurred is not enough; we also need to show that these changes were a direct consequence of the public health intervention rather than a result of other events or processes.

The Observed Change Occurred as a Consequence of the Intervention, Not Some Other Unrelated Event

If change has occurred in the behavior of interest, then a fundamental question is whether the change resulted from the intervention rather than from some other factor. This question can be answered in RCTs examining the effectiveness of drug treatments, where only individuals assigned to receive the intervention can access the drug being tested. Owing to the elimination of selection bias through randomization, any intervention effect can be attributed to the intervention.

Behavioral interventions are inherently more difficult to rigorously test. Interventions are usually made up of multiple interacting components (13), and investigators need to ensure that the intervention being tested does not contaminate individuals who have been assigned to the control arm. For example, in testing the effectiveness of a dietary advice intervention delivered by primary care physicians, a design where individuals are randomly allocated to receive or not receive advice

may result in contamination between the intervention and control conditions. The same physician delivering both the intervention and the control treatment could inadvertently apply elements of the experimental intervention to patients in the control group. To overcome this, it may be preferable to select primary care practices rather than individual patients and then randomly allocate practices into experimental and control conditions. Under optimal circumstances, the primary care practices would be separated by distance and not professionally linked. Such a design would minimize the potential for contamination between intervention and control practices. However, ensuring that geographical and professional separation conditions are met presents inherent logistical difficulties, as does recruiting a sufficient number of primary care practices to ensure adequate statistical power to detect a meaningful difference in dietary outcomes.

Demonstrating that a change is linked to an intervention becomes even more difficult at a community level. New policies may be introduced by government (such as increased taxes or an effective mass media campaign) at the same time that the intervention is implemented. As a consequence, the control group may be impacted by these community-wide strategies, producing change in both the experimental and the control groups. In these circumstances it is difficult to attribute any observed change to the implemented strategies rather than to other uncontrolled community-wide factors.

One example of the difficulties in attributing demonstrated changes to a specific intervention is provided by an examination of the effect of random breath testing (RBT) in reducing road traffic deaths in Australia. When RBT was introduced in New South Wales (NSW) in 1982, a decline in road fatalities of 48% followed (42). However, there had been a preexisting gradual decline in road trauma in many Australian states since the 1970s. The total magnitude of the change immediately following the introduction of RBT, and the fact that a similar pattern of change was observed in other states following the introduction of RBT, suggested that the changes in fatalities were attributable to RBT rather than to other factors.

The Amount of Change That Occurs as a Consequence of the Intervention Is Perceived as Significant By Stakeholders

It is necessary to determine not only whether change has occurred but also whether the change is significant. Statistical definitions of significance relate to whether the difference observed due to an intervention is likely to be real or due to chance. Clinical significance relates to the size of the effect and its likely impact on the health of populations or individuals (58). Stakeholders such as policy makers, clinicians, and administrators may place major emphasis on factors such as numbers needed to treat, risks of harm, and overall costs associated with intervention delivery. Patients and families directly affected are likely to have a different perspective on the costs and benefits of an intervention. Determining how big of an effect is needed involves some subjectivity and is likely to vary depending on the perspectives of stakeholders. Although health economics appears to offer a potentially standardized method of examining differential cost-effectiveness of strategies, the weights and benefits may not always be perceived in the same way by all stakeholders.

Furthermore, the decision to upscale an effective intervention or treatment into practice may not be based only on the strength of evidence for the intervention. Decision makers such as politicians will be influenced by the perceptions of health advisors and consumer groups and by the acceptability of the decision amid the current political climate. A study examining different perceptions of the harm caused by various drugs (81) serves to illustrate this point. Politicians and community members both perceived that so-called hard drugs such as heroin and ice imposed the largest burden on their community, whereas health care workers nominated smoking and alcohol as representing the largest health burden (81).

Table 1 Decision points at which an alternative to the randomized controlled trial (RCT) (with randomization by individuals) may be considered

Decision points	Description of barrier to RCT	Potential alternative designs
Is it appropriate for control individuals to not receive the intervention?	A control group not appropriate owing to: ■ ethical reasons ■ not acceptable to target population	Interventions in which all participants are able to receive the intervention include: ■ Stepped wedge cluster RCT ■ Multiple baseline ■ Interrupted time series
Is contamination likely if individuals are randomized?	Random assignment of individuals within settings (e.g., clinic, hospital, community) may result in contamination between experimental arms	Cluster RCT where the unit of randomization is the clinic, hospital, community, etc. may overcome this problem.
Is the unit of analysis large (e.g., community)?	If the unit of analysis is large and there are constraints regarding: ■ availability of an adequate sample size ■ expense/feasibility of recruiting an adequate sample size	Designs that can involve fewer units include: ■ Stepped wedge cluster RCT ■ Multiple baseline ■ Interrupted time series ■ Controlled prepost
Is long-term follow-up needed?	If: ■ long-term follow-up is needed AND ■ it is likely to be impractical to maintain randomization (i.e., control group is likely to be exposed to the intervention over time)	Designs that do not require long-term protection from contamination include: ■ Stepped wedge cluster RCT ■ Multiple baseline ■ Interrupted time series
Is external validity the primary concern of the research?	The intervention requires adaption: ■ across settings ■ across populations	Designs where intervention can vary across settings: ■ Stepped wedge ■ Multiple baseline ■ Interrupted time series
Is the intervention multileveled?	The intervention is: ■ complex AND ■ targets multiple levels (e.g., individual, environmental factors targeted for change)	Designs that suit multiple units of analysis include: ■ Cluster RCT ■ Stepped wedge cluster RCT ■ Multiple baseline

APPROPRIATE RESEARCH DESIGNS FOR EVALUATING COMPLEX INTERVENTIONS

Organizations such as the British Medical Research Council have produced guidance relating to complex interventions, providing a useful framework for designing such evaluations (19). This document reinforces the view that even though experimental designs are preferable, they are not always practicable (19). Several alternate designs to the individual RCT can be useful for evaluating public health interventions (19, 63, 80). In **Table 1** we present a set of criteria to assess barriers to the RCT and to determine whether an alternative design is appropriate. We then provide brief examples of study designs that are alternatives to individual RCTs.

Cluster Randomized Controlled Trials

Cluster RCTs are a type of RCT in which groups rather than individuals are randomized (25, 26, 41, 67). The cluster RCT is particularly appropriate for evaluating public health interventions in which (*a*) there is a risk of contamination should individuals be randomized, (*b*) the nature of

the intervention is such that it is targeted toward a group of individuals (as is often a feature of public health interventions), or (*c*) it is logistically more feasible to deliver the intervention at a group level rather than at an individual level. The cluster RCT is a true experimental study and thus has the methodological rigor of an RCT design (41). However, cluster RCTs have particular analytical challenges given that individuals within groups or clusters are likely to be correlated and standard statistical methods assume that observations are independent. This correlation can result in confidence intervals that are narrower than they should be and may increase the probability of a type I error if appropriate methods of analysis are not utilized. Analyses can be undertaken on cluster-level summary measures or using individual-level data. Most statistical packages have procedures that allow investigators to analyze correlated data using methods such as random effects or generalized estimating equations (GEE).

Other challenges of cluster RCTs include recruiting an adequate number of clusters, recruiting an adequate number of individuals within clusters, determining an appropriate sample size a priori, and a potential imbalance if the number of clusters is small. Stratified randomization can be used to improve balance between groups where there are a small number of clusters (25, 41, 55). Furthermore, for a small number of clusters, even robust variance is underestimated, and investigators need to use alternative approaches, for example the jackknife method, to obtain appropriate variance estimates.

In addition, cluster-randomized trials are often very costly to design and implement (90). The complexity associated with diversity of ethical requirements across clusters can significantly impede progress (91), and the need for extensive negotiation with whole cultural groups regarding consent for randomization (18) can make a study impractical. A key concern for the cluster RCT design is randomization of some clusters to the control condition, particularly where withholding potential help from vulnerable groups for long periods is considered unethical.

Middleton and colleagues undertook a cluster randomized controlled trial to evaluate the effect of a multidisciplinary intervention to improve management of stroke patients admitted to acute stroke units in NSW, Australia (65, 66). The intervention targeted staff within the stroke units and involved implementation of new patient treatment protocols and a change in the way in which patients were managed within the clinical environment. Therefore, it was not possible to randomize individuals within stroke units to receive or not receive this intervention, and a cluster RCT was the most appropriate evaluation design.

Other examples of studies employing a cluster RCT design include the Cancer Action in Rural Towns (CART) project (35–38) and the Alcohol Action in Rural Communities (AARC) project (85). The CART project aimed to evaluate a community action program for increasing rates of preventive and screening behaviors relating to breast, cervical, smoking-related, and skin cancers (38). The design involved 20 towns in rural NSW, selected using a matched-pair design, with one town from each pair randomly allocated to the experimental or control condition. In relation to smoking outcomes, results of the project indicated that 40 towns would have been required to show a significant difference of 3.5% in quit rates (37). This requirement illustrates just one of the challenges involved in employing the cluster RCT design (37).

The AARC project also used a matched-pair cluster RCT design to evaluate the effectiveness of a community action strategy aimed at reducing alcohol misuse and alcohol-related harm (85). The project was one of the largest evaluations of the community action approach to reducing alcohol problems ever undertaken (86), involving 20 communities as the unit of randomization and analysis. The challenges in undertaking this large community comparison study included the high cost, the level of expertise needed, and the long-term commitment required from researchers and community partners, which included local government, state and federal agencies, not-for-profit organizations, and businesses (86).
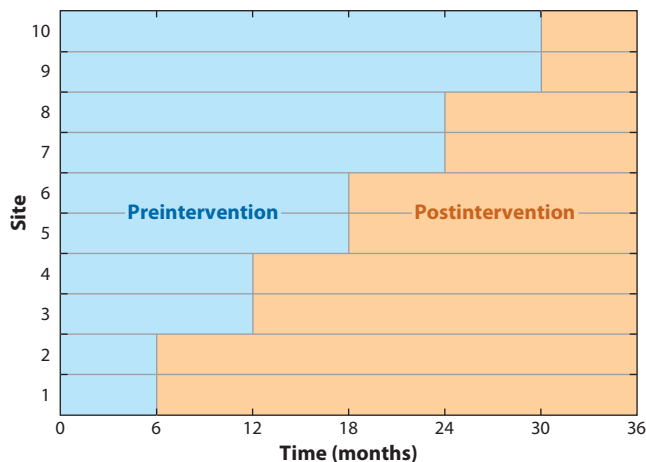
Figure 1

Stepped wedge design in which the intervention is implemented sequentially in all clusters or sites and is delivered to two sites simultaneously.

Stepped Wedge Design

A modification of the cluster RCT is the stepped wedge design. This design is fairly new and has been gaining popularity (12), although some still have concerns about its utility (56, 57, 61, 62). In a stepped wedge design, the intervention is implemented sequentially in all clusters. The concept is similar to a crossover RCT, but it is unidirectional so that at the start of the trial all clusters are in the control phase and by the end of the study all clusters are in the intervention phase (**Figure 1**). The order in which sites receive the intervention is randomized, and thus the study is an experimental design. The intervention may be implemented in multiple sites at the same time, as shown in **Figure 1** where the intervention is delivered to two sites simultaneously. This study design is a good alternative to the standard cluster RCT if it is important that all clusters receive the intervention (e.g., Aboriginal populations) but it is not possible, for logistical, practical, or financial reasons, for investigators to deliver the intervention simultaneously to all participants (11).

In a stepped wedge design, the outcomes are measured at the start of the study prior to implementing the intervention in any of the sites then measured again at the end of each intervention period and before implementation in the next site/s. If implementation requires a long time period, additional measures may be taken during this period. The design may thus require more measurements than a standard cluster RCT. The stepped wedge design can be more efficient than the cluster RCT under some conditions; in other situations it may have less power and require larger sample sizes. The design can involve cohort or repeated cross-sectional data. The former can, however, introduce temporal bias (11, 48).

Design and analysis issues are generally similar to those of a standard cluster RCT, but with additional complexity. Data can be analyzed vertically or horizontally to control for temporal trends. This process can involve obtaining a weighted average of between-group differences at each measurement time across all times. This calculation assumes a constant intervention effect, with effective "blocking" by time period. The difference between intervention and control clusters is obtained at each measurement time, and then a weighted average over all time periods is calculated (with weights inversely proportional to variance). Alternatively, a weighted average of within-group differences over time across all groups can be obtained. If there is no time effect, investigators then compare pre- to posttest differences in each cluster, and a weighted average

of within-cluster differences is obtained. The within- and between-group information can be combined using a generalized linear model with both intervention-phase and time factors; this method of analysis is usually more efficient than between- or within-group comparison only.

Advantages of the stepped wedge cluster design over the parallel cluster RCT design include allowing all, rather than only half, of the clusters to receive the intervention; the intervention can also be introduced at staggered times, which may provide logistic/feasibility advantages under some conditions. Disadvantages include an increased number of measurements, resulting in a longer evaluation period because study power decreases with fewer measurement points; design and analysis are also more complex. Variation in cluster size can reduce the efficiency of a parallel cluster RCT. Work that evaluates the value of these designs, in particular the number of clusters, the number of observations per cluster, and the number of measurements required, as well as the most appropriate method of analysis, is still limited.

An early example of a stepped wedge design was the Gambia Hepatitis Intervention Study (31), which commenced in 1986 and aimed to investigate the long-term (30-year) effect of hepatitis B virus vaccination for infants on liver cancer and other chronic liver diseases. A new group of children was vaccinated every three months over the four-year study intervention period. The stepped wedge design was optimal for logistical reasons related to vaccine availability and cost and the feasibility of vaccinating the large number of children within a short time frame, necessitating a phased approach to implementation. More recently, investigators used a stepped wedge design to evaluate whether integrating antiretroviral therapy (ART) into antenatal care in Zambia increased the proportion of women with HIV who initiate ART therapy during pregnancy compared with the standard referral approach (53). The intervention was implemented in eight antenatal clinics using a stepped approach. The stepped wedge design was considered an appropriate choice in this case because withholding this treatment would have been considered unethical.

Interrupted Time Series Design

The interrupted time series design is a method often used for evaluating an intervention delivered to a single site, group, or cluster. Multiple measures of a cluster-specific summary variable are obtained pre- and postintervention; thus the site acts as its own control (see **Figure 2**). The analysis involves fitting an interrupted time series to the data. A time series is fit pre- and postintervention using a disjointed segmented linear regression model to assess whether there is a change in intercept and/or slope from pre- to postintervention periods. A change in intercept indicates a jump or change in outcome following the intervention, whereas a change in slope indicates a different trend in outcome postintervention. Statistical analysis can involve individual assessment of prepost intervention changes in intercept or slope or a global test of both parameters (20). Analyses need to adjust for (serial) correlation among measures using an autoregressive term.

The interrupted times series design is particularly relevant where the outcome is obtained from routinely collected data, for example hospital admissions, crime statistics, and medication use, and is useful for evaluating policy change. Methodological issues include determining the number of observations (time periods) required and the duration of the time period of measurement; for example, using monthly summary data rather than quarterly data will provide more measures but less precision per time period. The major disadvantage of the interrupted time series design is that investigators cannot determine with confidence whether any change from pre- to postintervention is due to the intervention alone and not to external factors. The interrupted time series is classified as a quasi-experimental design because there is no random allocation of treatment (74, 84).

An interesting application of the interrupted time series approach was the evaluation of the impact of changes in firearms legislation implemented in Australia in 1988 and 1996, following
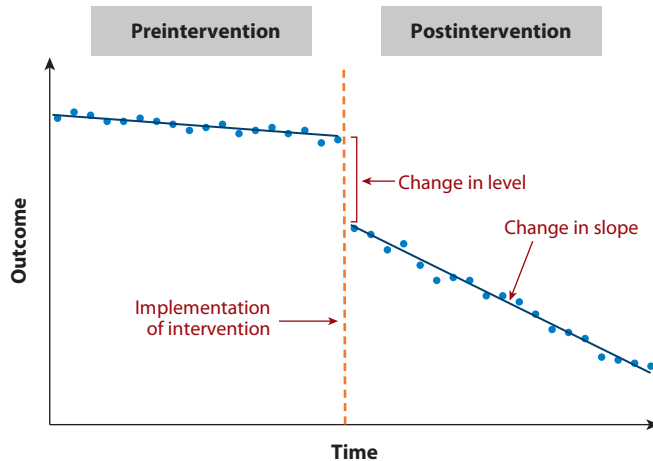
Figure 2

Interrupted time series design in which a change from pre- to postintervention outcome indicates an intervention effect; the change in level indicates a change in outcome, whereas the change in slope indicates a different trend in outcome following the intervention.

two mass shooting incidents in which 15 people were killed in Victoria, in 1987, and 35 people were killed in Port Arthur, Tasmania, in 1996 (69). This study investigated the annual rates of firearm-related deaths over time and showed a decline in this outcome following each of the legislation changes.

Multiple Baseline Design

The multiple baseline design offers a solution to the main disadvantage of the interrupted time series design noted above: namely, lack of confidence that any effect is due to the intervention rather than to extraneous factors (40). The multiple baseline design is similar to the interrupted time series design but uses multiple sites, as shown in **Figure 3**. The intervention is implemented in a phased approached to the multiple sites at different times, with the time of intervention implementation randomized. Demonstration of an intervention effect in one site but not another (at the same time), with repetition of the intervention effect after implementation of the intervention in each site, reduces the likelihood that any effect is due to external factors. A combined analysis involves a generalized linear model using intervention condition, time, and site, adjusted for multiple observations per site (9). Because the multiple baseline design involves observations pre- and postintervention in all sites or groups, it can also be considered a stepped wedge design. Stepped wedge design usually refers to cluster RCTs (i.e., multiple individual measures per unit per time), whereas multiple baseline design usually refers to a single measurement per unit per time (e.g., number of prescriptions, accidents). Because they involve random allocation of treatment order, multiple baseline studies are also experimental designs (74, 84).

Biglan and colleagues (8, 10) used a multiple baseline study design to investigate the effect of a multifaceted community intervention to reduce tobacco sales to young people. The intervention involved generating community support from a range of government organizations, health care providers, schools, churches, social welfare agencies, businesses, and tobacco outlets. The study included four communities, arranged in two pairs; one community from each pair
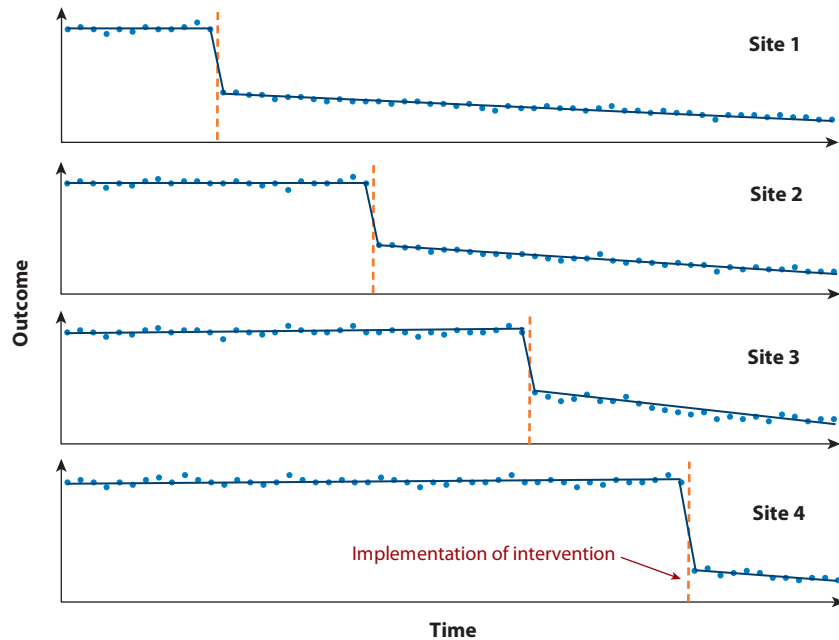
Figure 3

Multiple baseline design in which the intervention is implemented in a phased approach in four different sites at different times.

initially received the intervention, then the intervention was delivered in the second community approximately eight months later. The study was then replicated in four more communities. Analysis involved undertaking interrupted time series analysis in the four intervention communities.

Controlled Prepost Designs

The interrupted time series and multiple baseline designs are both forms of before and after (or prepost) study designs. An alternative, less rigorous type of study is an uncontrolled prepost design involving only one observation, rather than multiple observations, before and after the intervention. This type of study is easy to undertake; however, the major flaw is that any effect cannot be ascribed to the intervention because there is potential for influence from extraneous factors and temporal trends. The addition of a nonintervention site to form a controlled prepost study improves the validity of this design (30). This control site may be selected randomly to produce an experimental design or for convenience (a less ideal, but often more feasible choice and a quasi-experimental design). A change in outcome(s) from preintervention to postintervention in the intervention site but not the control site provides evidence that the effect is due to the intervention. For example, the impact of the implementation of a hospital guideline for management of patients with community-acquired pneumonia was compared between an intervention hospital and four randomly selected control hospitals (14). Data on processes of care and outcomes for patients with community-acquired pneumonia were abstracted from medical records for 12 months prior to the guideline implementation and 19 months after its implementation. Data were obtained from control hospitals for the same time periods.

Table 2 Advantages and disadvantages of alternative research designs for complex interventions

Study design	Advantages	Disadvantages	Ability to answer key methodological questions		
			Has change occurred?	Was the change due to the intervention?	Is the degree of change significant?
Cluster randomized controlled trial (RCT)	■ Allows randomisation of groups or clusters, reducing selection bias ■ Appropriate when intervention is targeted toward a group rather than to individuals ■ Reduces risk of contamination ■ May be logistically more feasible to deliver intervention at group rather than individual level	■ Analysis needs to adjust for correlations within clusters ■ Generally requires larger numbers of individuals ■ Difficulties arise in recruiting an adequate number of clusters and number of individuals per cluster ■ Small number of clusters may result in either covariate imbalance between groups or underestimation of variance ■ Requires a control condition that may not be acceptable in some situations ■ Information is limited on the likely degree of correlation estimated by the intraclass correlation coefficient (ICC) or coefficient of variation for sample size estimation	Strong[a]	Strong[a]	Strong[a]
Stepped wedge design	■ Allows randomization of order of receipt of intervention, thus reducing selection bias ■ All clusters receive the intervention ■ Staggering of implementation can be more feasible/efficient than a standard cluster RCT design	■ Requires that all clusters receive the intervention (sequentially) ■ Generally requires more measurements and a longer evaluation period than a standard cluster RCT design ■ Can be less powerful and require larger sample sizes than a cluster RCT design ■ Data analysis may be more complex than for a cluster RCT	Strong[b]	Moderate[c]	Moderate
Interrupted time series (ITS)	■ Only requires a single group or cluster ■ No need to allocate clusters/sites to a control condition ■ Can utilize routinely collected data such as hospital admissions data or crime statistics	■ Requires a large number of measurements and long data collection period ■ Limited confidence that results are due to the intervention rather than to external factors	Moderate	Weak	Moderate

(Continued)

Table 2   (*Continued*)

Study design	Advantages	Disadvantages	Ability to answer key methodological questions		
			Has change occurred?	Was the change due to the intervention?	Is the degree of change significant?
Multiple baseline	■ Allows randomization of clusters to order of intervention, thus reducing selection bias ■ Provides more confidence than an ITS design that change is due to the intervention ■ All clusters receive intervention	■ Requires multiple clusters and that all clusters receive the intervention (sequentially) ■ May require more measurements than ITS (for multiple sites) and a longer evaluation period than a cluster RCT design ■ Data analysis may be more complex than for a cluster RCT	Moderate	Moderate	Moderate
Controlled prepost	■ Design and implementation are simplified ■ Control site can be selected randomly or for convenience	■ Confidence that results are due to the intervention rather than to temporal trends and/or external factors is limited ■ Generally small number of clusters so limited generalizability	Moderate	Weak[d]	Moderate

[a] Statistical methods that allow adjustment for clustering and/or adjustment for small numbers of clusters are required.
[b] May have less power than the RCT to detect a change.
[c] Results may also be subject to temporal trends.
[d] Addition of a nonintervention site to form a controlled prepost study improves validity.

Although none of the designs described above provides a simple and complete solution to the challenges of evaluating real-world public health interventions (see **Table 2**), it is clear that a range of options is available, along with an increasing sophistication in our understanding of these designs.

SUMMARY AND CONCLUSIONS

Public health interventions must be rigorously evaluated to answer the key questions of whether a positive change has occurred, whether any observed change occurred as a result of the intervention of interest, and whether the degree of change was significant. Many public health problems require complex interventions that acknowledge multiple influences on behavior. This complexity may necessitate intervening with whole systems rather than with individuals. A range of alternatives to the individual RCT can be used for evaluating such interventions, including the cluster RCT, stepped wedge, interrupted time series, multiple baseline, and controlled prepost designs. The choice of design needs to take into account factors such as the complexity of the intervention, the number of units or systems that can be recruited, the cost of evaluation, the costs and logistical implications of obtaining repeated measures, and ethical issues related to withholding the intervention from some groups.

SUMMARY POINTS

1. Public health interventions are often complex and multifaceted and target systems rather than individuals.

2. Evaluation of public health interventions is essential to ensure efficient use of scarce resources, maximize health benefits, minimize harms, and avoid exacerbation of health inequities.

3. The traditional gold standard approach to intervention evaluation, the individually randomized controlled trial, may be incompatible with the need to intervene with systems and target multiple factors that influence public health behavior.

4. Alternative research designs, such as the cluster RCT, stepped wedge, interrupted time series, multiple baseline, and controlled prepost designs, each offer advantages and complexities as compared with the RCT.

5. The key features, advantages, and complexities of these alternative designs should be taken into account when choosing a study design.

DISCLOSURE STATEMENT

LITERATURE CITED

1. Adler A. 1992. *What Life Could Mean to You*. Oxford, UK: Oneworld

2. Ajzen I. 1985. From intentions to actions: a theory of planned behaviour. In *Action Control: From Cognition to Behavior*, ed. J Kuhl, J Beckmann, pp. 11–40. Berlin: Springer-Verlag

3. Aust. Bur. Stat. 2006. *Australian social trends, 2000, 4102.0*. Last updated May 2. **http://www.abs.gov.au/ausstats/abs@.nsf/2f762f95845417aeca25706c00834efa/738d23457b86defbca2570ec000e2f5c!OpenDocument**

4. Aust. Inst. Health Welf. 2011. *2010 National Drug Strategy Household Survey Report. Drug Stat. Ser. No. 25*. Cat. no. PHE 145. Canberra: Aust. Inst. Health Welf. **http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=10737421314**

5. Aust. Inst. Health Welf. 2013. *Public Health Expenditure in Australia 2008–2009*. Canberra: Aust. Inst. Health Welf. **http://www.aihw.gov.au/public-health-expenditure/**

6. Bandura A. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall

7. Baum F. 1998. *The New Public Health: An Australian Perspective*. Melbourne, Aust.: Oxford Univ. Press

8. Biglan A, Ary D, Koehn V, Levings D, Smith S, et al. 1996. Mobilizing positive reinforcement in communities to reduce youth access to tobacco. *Am. J. Community Psychol.* 24:625–38

9. Biglan A, Ary D, Wagenaar AC. 2000. The value of interrupted time-series experiments for community intervention research. *Prev. Sci.* 1:31–49

10. Biglan A, Henderson J, Humphrey D, Yasui M, Whisman R, et al. 1995. Mobilising positive reinforcement to reduce youth access to tobacco. *Tob. Control* 4:42–48

11. Brown CA, Lilford RJ. 2006. The stepped wedge trial design: a systematic review. *BMC Med. Res. Methodol.* 6:54–63

12. Brown CH, Wyman PA, Guo J, Peña J. 2006. Dynamic wait-listed designs for randomized trials: new designs for prevention of youth suicide. *Clin. Trials* 3:259–71

13. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, et al. 2000. Framework for design and evaluation of complex interventions to improve health. *BMJ* 321:694–96

14. Capelastegui A, España PP, Quintana JM, Gorordo I, Ortega M, et al. 2004. Improvement of process-of-care and outcomes after implementing a guideline for the management of community-acquired pneumonia: a controlled before-and-after design study. *Clin. Infect. Dis.* 39:955–63

15. CDC (Cent. Dis. Control Prev.). 2011. Vital signs: current cigarette smoking among adults aged ≥18 years—United States, 2005–2010. *MMWR* 60:1207–12

16. CDC (Cent. Dis. Control Prev.). 2012. *Trends in current cigarette smoking among high school students and adults, United States, 1965–2011.* Updated November 8, 2013. **http://www.cdc.gov/tobacco/data_statistics/tables/trends/cig_smoking/index.htm**

17. Cent. Evidence Based Medicine. 2013. *Oxford Centre for Evidence-based Medicine—levels of evidence* (*March 2009*). Updated September 16. **http://www.cebm.net/?o=1025**

18. Corrigan O. 2003. Empty ethics: the problem with informed consent. *Sociol. Health Illn.* 25:768–92

19. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. 2008. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 337:a1655

20. Crosbie J. 1993. Interrupted time-series analysis with brief single-subject data. *J. Consult. Clin. Psychol.* 61:966–74

21. DeVita VT Jr. 2005. The Framework Convention on Tobacco Control. *Nat. Clin. Pract. Oncol.* 2:171

22. Dickinson J, Wiggers J, Leeder S, Sanson-Fisher RW. 1989. General practitioners' detection of patients' smoking status. *Med. J. Aust.* 150:420–22, 425–26

23. Djulbegovic M, Beyth RJ, Neuberger MM, Stoffs TL, Vieweg J, et al. 2010. Screening for prostate cancer: systematic review and meta-analysis of randomised controlled trials. *BMJ* 341:c4543

24. Donabedian A. 1988. The quality of health care: How can it be assessed? *JAMA* 260:1743–48

25. Donner A, Klar N. 2000. *Design and Analysis of Cluster Randomization Trials in Health Research*. London: Arnold Publ.

26. Eldridge S, Kerry S. 2012. *A Practical Guide to Cluster Randomised Trials in Health Services Research*. Chichester, UK: Wiley

27. Erbas B, Bui Q, Huggins R, Harper T, White V. 2006. Investigating the relation between placement of Quit antismoking advertisements and number of telephone calls to Quitline: a semiparametric modelling approach. *J. Epidemiol. Community Health* 60:180–82

28. Ezzati M, Friedman AB, Kulkarni SC, Murray CJ. 2008. The reversal of fortunes: trends in county mortality and cross-county mortality disparities in the United States. *PLoS Med.* 5:e66

29. Freud S. 1974. *The Standard Edition of the Complete Psychological Works of Sigmund Freud:* Vol. 11. Great Britain: Hogarth

30. Freund M, Campbell E, Paul C, Sakrouge R, Lecathelinais C, et al. 2009. Increasing hospital-wide delivery of smoking cessation care for nicotine-dependent in-patients: a multi-strategic intervention trial. *Addiction* 104:839–49

31. Gambia Hepatitis Study Group. 1987. The Gambia Hepatitis Intervention Study. *Cancer Res.* 47:5782–87

32. Glasziou P, Vandenbroucke J, Chalmers I. 2004. Assessing the quality of research. *BMJ* 328:39–41

33. Green LW, Kreuter MW. 1991. *Health Promotion Planning: An Educational and Environmental Approach*. Mountain View, CA: Mayfield. 2nd ed.

34. Han PK, Coates RJ, Uhler RJ, Breen N. 2006. Decision making in prostate-specific antigen screening: National Health Interview Survey, 2000. *Am. J. Prev. Med.* 30:394–404

35. Hancock L, Sanson-Fisher R, Perkins J, Corkrey R, Burton R, Reid S. 2001. Effect of a community action intervention on cervical cancer screening rates in rural Australian towns: the CART Project. *Prev. Med.* 32:109–17

36. Hancock L, Sanson-Fisher R, Perkins J, Girgis A, Howley P, Schofield M. 2001. The effect of a community action intervention on adolescent smoking rates in rural Australian towns: the CART project. *Prev. Med.* 32:332–40

37. Hancock L, Sanson-Fisher R, Perkins J, McClintock A, Howley P, Gibberd R. 2001. Effect of a community action program on adult quit smoking rates in rural Australian towns: the CART Project. *Prev. Med.* 32:118–27

38. Hancock L, Sanson-Fisher R, Redman S, Burton R, Burton L, et al. 1996. Community action for cancer prevention: overview of the Cancer Action in Rural Towns (CART) project, Australia. *Health Prom. Int.* 11:277–90

39. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, et al. 2001. Current methods of the U.S. Preventive Services Task Force: a review of the process. *Am. J. Prev. Med.* 20:21–35

40. Hawkins N, Sanson-Fisher RW, Shakeshaft A, D'Este C, Green LW. 2007. The multiple baseline design for evaluating population-based research. *Am. J. Prev. Med.* 33:162–68

41. Hayes RJ, Moulton LH. 2009. *Cluster Randomised Trials*. London: CRC Press

42. Henstridge J, Homel R, Mackay P. 1997. *The Long-Term Effects of Random Breath Testing in Four Australian States: A Time Series Analysis*. CR 162. Canberra, Aust.: Fed. Off. Road Saf.

43. Heywood A, Sanson-Fisher R, Ring I, Mudge P. 1994. Risk prevalence and screening for cancer by general practitioners. *Prev. Med.* 23:152–59

44. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, et al. 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 343:d5928

45. Higgins JPT, Green S, eds. 2009. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.0.2, updated September. Chichester, UK: Wiley

46. Hillestad R, Bigelow J, Bower A, Girosi F, Meili R, et al. 2005. Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Aff.* 24:1103–17

47. Horrocks S, Anderson E, Salisbury C. 2002. Systematic review of whether nurse practitioners working in primary care can provide equivalent care to doctors. *BMJ* 324:819–23

48. Hussey MA, Hughes JP. 2007. Design and analysis of stepped wedge cluster randomized trials. *Contemp. Clin. Trials* 28:182–91

49. Ilic D, O'Connor D, Green S, Wilt TJ. 2006. Screening for prostate cancer. *Cochrane Database Syst. Rev.* 1:CD004720

50. Janz NK, Becker MH. 1984. The health belief model: a decade later. *Health Educ. Behav.* 11:1–47

51. Jeffreys M, Stevanovic V, Tobias M, Lewis C, Ellison-Loschmann L, et al. 2005. Ethnic inequalities in cancer survival in New Zealand: linkage study. *Am. J. Public Health* 95:834–37

52. Katsuta T, Saitoh A. 2012. Haemophilus influenzae type b vaccine. *J. Infect. Dis.* 206:1645–51

53. 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* 24:85–91

54. Kitchen MS, Ransley JK, Greenwood DC, Clarke GP, Conner MT, et al. 2009. Study protocol: a cluster randomised controlled trial of a school based fruit and vegetable intervention—Project Tomato. *BMC Health Serv. Res.* 9:101

55. Klar N, Donner A. 1997. The merits of matching in community intervention trials: a cautionary tale. *Stat. Med.* 16:1753–64

56. Kotz D, Spigt M, Arts IC, Crutzen R, Viechtbauer W. 2012. Researchers should convince policy makers to perform a classic cluster randomized controlled trial instead of a stepped wedge design when an intervention is rolled out. *J. Clin. Epidemiol.* 65:1255–56

57. Kotz D, Spigt M, Arts IC, Crutzen R, Viechtbauer W. 2012. Use of the stepped wedge design cannot be recommended: a critical appraisal and comparison with the classic cluster randomized controlled trial design. *J. Clin. Epidemiol.* 65:1249–52

58. Leung WC. 2001. Balancing statistical and clinical significance in evaluating treatment effects. *Postgrad. Med. J.* 77:201–4

59. Macintyre S. 2003. Evidence based policy making: Impact on health inequalities still needs to be assessed. *BMJ* 326:5–6

60. Marmot M. 2005. Social determinants of health inequalities. *Lancet* 365:1099–104

61. Mdege ND, Man MS, Taylor (nee Brown) CA, Torgerson DJ. 2012. There are some circumstances where the stepped-wedge cluster randomized trial is preferable to the alternative: no randomized trial at all. Response to the commentary by Kotz and colleagues. *J. Clin. Epidemiol.* 65:1253–54

62. Mdege ND, Man MS, 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

63. Mercer S, DeVinney B, Fine L, Green L, Dougherty D. 2007. Study designs for effectiveness and translation research: identifying trade-offs. *Am. J. Prev. Med.* 33:139–54
64. Michie S, van Stralen MM, West R. 2011. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement. Sci.* 6:42–53
65. Middleton S, Levi C, Ward J, Grimshaw J, Griffiths R, et al. 2009. Fever, hyperglycaemia and swallowing dysfunction management in acute stroke: a cluster randomised controlled trial of knowledge transfer. *Implement. Sci.* 4:16–27
66. Middleton S, McElduff P, Ward J, Grimshaw J, Dale S, et al. 2011. Implementation of evidence-based treatment protocols to manage fever, hyperglycaemia, and swallowing dysfunction in acute stroke (QASC): a cluster randomised controlled trial. *Lancet* 378:1699–706
67. Murray DM. 1998. *Design and Analysis of Group-Randomized Trials*. New York: Oxford Univ. Press
68. Nestle M. 2005. Preventing childhood diabetes: the need for public health intervention. *Am. J. Public Health* 95:1497–99
69. Ozanne-Smith J, Ashby K, Newstead S, Stathakis V, Clapperton A. 2004. Firearm related deaths: the impact of regulatory reform. *Inj. Prev.* 10:280–86
70. Pan SY, Cameron C, DesMeules M, Morrison H, Craig CL, Jiang X. 2009. Individual, social, environmental, and physical environmental correlates with physical activity among Canadians: a cross-sectional study. *BMC Public Health* 9:21–33
71. Pawson R, Tilley N. 2006. Realist evaluation. *DPRN Them. Meet. 2006 Rep. Eval.* Amsterdam: Dev. Policy Rev. Netw. (DPRN). **http://www.communitymatters.com.au/RE_chapter.pdf**
72. Petticrew M, Whitehead M, Macintyre S, Graham H, Egan M. 2004. Evidence for public health policy on inequalities: 1: The reality according to policymakers. *J. Epidemiol. Community Health* 58:811–16
73. Reason P, Bradbury H, eds. 2008. *The Sage Handbook of Action Research: Participative Inquiry and Practice*. Thousand Oaks, CA: Sage
74. Reichardt CS, Mark MM. 2004. Quasi-experimentation. In *Handbook of Practical Program Evaluation*, ed. HH Wholey, JK Newcomer, pp. 126–49. San Francisco, CA: Jossey-Bass Inc.
75. Reid AL, Webb GR, Hennrikus D, Fahey PP, Sanson-Fisher RW. 1986. Detection of patients with high alcohol intake by general practitioners. *BMJ Clin. Res.* 293:735–37
76. Reid JL, Hammond D, Boudreau C, Fong GT, Siahpush M. 2010. Socioeconomic disparities in quit intentions, quit attempts, and smoking abstinence among smokers in four western countries: findings from the International Tobacco Control Four Country Survey. *Nicotine Tob. Res.* 12:S20–33
77. Rogers CR, Skinner BF. 1956. Some issues concerning the control of human behaviour. *Science* 124:1057–66
78. Rosenthal MB, Frank RG, Li Z, Epstein AM. 2005. Early experience with pay-for-performance: from concept to practice. *JAMA* 294:1788–93
79. Rychetnik L, Frommer M, Hawe P, Shiell A. 2002. Criteria for evaluating evidence on public health interventions. *J. Epidemiol. Community Health* 56:119–27
80. Sanson-Fisher R, Bonevski B, Green L, D'Este C. 2007. Limitations of the randomised controlled trial in evaluating population-based health interventions. *Am. J. Prev. Med.* 33:155–61
81. Sanson-Fisher RW, Redman S, Oak S, Webb G. 1988. The burden of illness that is imposed by drug abuse and the allocation of research monies in the field: three views. *Med. J. Aust.* 149:134–38
82. Saunders RP, Evans MH, Joshi P. 2005. Developing a process-evaluation plan for assessing health promotion program implementation: a how-to guide. *Health Promot. Pract.* 6:134–47
83. Schaap MM, Kunst AE, Leinsalu M, Regidor E, Ekholm O, et al. 2008. Effect of nationwide tobacco control policies on smoking cessation in high and low educated groups in 18 European countries. *Tob. Control* 17:248–55
84. Shadish WR, Cook TD, Campbell DT. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston/New York: Houghton Mifflin
85. Shakeshaft A, Doran C, Petrie D, Breen C, Havard A, et al. 2012. *The Alcohol Action in Rural Communities (AARC) Project: Detailed Description of the Interventions and Their Costs*. Sydney: Natl. Drug Alcohol Res. Cent., Univ. N.S.W.

86. Shakeshaft A, Doran C, Petrie D, Breen C, Havard A, et al. 2012. *The Alcohol Action in Rural Communities (AARC) Project: Working with Communities to Select, Implement and Measure the Impact of Strategies to Reduce Alcohol-Related Harm.* Deakin, Aust.: Found. Alcohol Res. Educ.

87. Shen D, Lu Z. 2006. *Randomization in clinical trial studies.* Pap. Po06 presented at PharmaSUG, May 21–24, Bonita Springs, Fla. **http://www.lexjansen.com/pharmasug/2006/posters/po06.pdf**

88. Swan J, Breen N, Coates RJ, Rimer BK, Lee NC. 2003. Progress in cancer screening practices in the United States: results from the 2000 National Health Interview Survey. *Cancer* 97:1528–40

89. Tabrizi SN, Brotherton JM, Kaldor JM, Skinner SR, Cummins E, et al. 2012. Fall in human papillo-mavirus prevalence following a national vaccination program. *J. Infect. Dis.* 206:1645–51

90. Thompson B, Coronado G, Snipes SA, Puschel K. 2003. Methodologic advances and ongoing challenges in designing community-based health promotion programs. *Annu. Rev. Public Health* 24:315–40

91. Tully J, Ninis N, Booy R, Viner R. 2000. The new system of review by multicentre research ethics committees: prospective study. *BMJ* 320:1179–82

92. Volmink J, Garner P. 2007. Directly observed therapy for treating tuberculosis. *Cochrane Database Syst. Rev.* (4):CD003343

93. Wakefield MA, Loken B, Hornik RC. 2010. Use of mass media campaigns to change health behaviour. *Lancet* 376:1261–71

94. Wang J, Etter JF. 2004. Administering an effective health intervention for smoking cessation online: the international users of Stop-Tabac. *Prev. Med.* 39:962–68

95. Wendy B. 2012. Vaccination with 3-dose paediatric rotavirus vaccine (RotaTeq): impact on the timeliness of uptake of the primary course of DTPa vaccine. *Vaccine* 30:5293–97

96. WHO (World Health Organ.). 2005. *Preventing Chronic Diseases: A Vital Investment.* Geneva: WHO

97. WHO (World Health Organ.). 2011. *Global Status Report on Noncommunicable Diseases 2010: Description of the Global Burden of NCDs, Their Risk Factors and Determinants.* Geneva: WHO

98. WHO Global Tuberc. Program. 1997. *WHO Report on the Tuberculosis Epidemic 1997.* Geneva: WHO

99. Wilbur J. 2008. Prostate cancer screening: the continuing controversy. *Am. Fam. Physician* 78:1377–84

100. Wolf A, Wender RC, Etzioni RB, Thompson IM, D'Amico AV, et al. 2010. American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA Cancer J. Clin.* 60:70–98

101. Zaza S, Briss PA, Harris KW, eds. 2005. *The Guide to Community Preventive Services: What Works to Promote Health?* New York: Oxford Univ. Press

102. Ziglio E. 1997. How to move towards evidence-based health promotion interventions. *Promot. Educ.* 4:29–33