



Published in final edited form as:

*Annu Rev Public Health*. 2017 March 20; 38: 1–22. doi:10.1146/annurev-publhealth-031816-044215.

## An Overview of Research and Evaluation Designs for Dissemination and Implementation

**C. Hendricks Brown<sup>1</sup>, Geoffrey Curran<sup>2</sup>, Lawrence A. Palinkas<sup>3</sup>, Gregory A. Aarons<sup>4</sup>, Kenneth B. Wells<sup>5</sup>, Loretta Jones<sup>6</sup>, Linda M. Collins<sup>7</sup>, Naihua Duan<sup>8</sup>, Brian S. Mittman<sup>9</sup>, Andrea Wallace<sup>10</sup>, Rachel G. Tabak<sup>11</sup>, Lori DuCharme<sup>12</sup>, David Chambers<sup>13</sup>, Gila Neta<sup>13</sup>, Tisha Wiley<sup>14</sup>, John Landsverk<sup>15</sup>, Ken Cheung<sup>16</sup>, and Gracelyn Cruden<sup>1</sup>**

<sup>1</sup>Department of Psychiatry and Behavioral Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611

<sup>2</sup>Division of Health Services Research, Psychiatric Research Institute, University of Arkansas for Medical Sciences, Little Rock, AR 72205

<sup>3</sup>Department of Children, Youth and Families, School of Social Work, University of Southern California, Los Angeles, CA 90089

<sup>4</sup>Department of Psychiatry, University of California, San Diego, School of Medicine, La Jolla, CA 92093

<sup>5</sup>Center for Health Services and Society, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA 90024

<sup>6</sup>Healthy African American Families, Los Angeles, CA USA

<sup>7</sup>The Methodology Center and Department of Human Development & Family Studies, Pennsylvania State University, University Park, PA 16802

<sup>8</sup>Department of Psychiatry, Columbia University Medical Center, Columbia University, New York, NY, 10027

<sup>9</sup>VA Center for Implementation Practice and Research Support, VA Greater Los Angeles Healthcare System, North Hills, CA 91343

<sup>10</sup>College of Nursing, The University of Iowa, Iowa City, IA 52242

<sup>11</sup>Prevention Research Center, George Warren Brown School, Washington University, St Louis, MO 63105

<sup>12</sup>National Institute of Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, MD 20814

<sup>13</sup>Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD 20850

<sup>14</sup>National Institute on Drug Abuse, National Institutes of Health, Bethesda, MD 20814

---

**Corresponding Author:** C. Hendricks Brown (hendricks.brown@northwestern.edu), (312)503-4324, 750 N Lake Shore Dr. 10<sup>th</sup> Floor, Chicago, IL 60611.

<sup>15</sup>Oregon Social Learning Center, Eugene, OR 97401

<sup>16</sup>Mailman School of Public Health, Columbia University, New York, NY 10032

## Abstract

**Background**—The wide variety of dissemination and implementation designs now being used to evaluate and improve health systems and outcomes warrants review of the scope, features, and limitations of these designs.

**Methods**—This paper is one product of a design workgroup formed in 2013 by the National Institutes of Health to address dissemination and implementation research, and whose members represented diverse methodologic backgrounds, content focus areas, and health sectors. These experts integrated their collective knowledge on dissemination and implementation designs with searches of published evaluations strategies.

**Results**—This paper emphasizes randomized and non-randomized designs for the traditional translational research continuum or pipeline, which builds on existing efficacy and effectiveness trials to examine how one or more evidence-based clinical/prevention interventions are adopted, scaled up, and sustained in community or service delivery systems. We also mention other designs, including hybrid designs that combine effectiveness and implementation research, quality improvement designs for local knowledge, and designs that use simulation modeling.

## Keywords

implementation trial; scale up; adoption; adaptation; fidelity; sustainment

---

## Introduction

Medicine and public health have made great progress through rigorous randomized clinical trials in determining whether an intervention is efficacious. The standards set by Fisher (39), who laid the foundation for experimental design first in agriculture, and by Hill (58), who developed the randomized clinical trial for medicine, provided a unified approach to examining the efficacy of an individual-level intervention versus control condition or comparative effectiveness of one active intervention against another. Practical modifications of the individual-level randomized clinical trial have been made to test for program or intervention effectiveness under realistic conditions in randomized field trials (40), as well as for interventions delivered at the group level (77), multilevel interventions (14, 15), complex (34, 70) or multiple component interventions (31), and tailoring or adapting the intervention to subject's response to targeted outcomes (31, 74) or to different social, physical, or virtual environments (13). There is now a large family of designs for testing efficacy or effectiveness that randomize across persons, place, and time (or combinations of these) (15), as well as designs that do not use randomization, such as pre-post comparisons, regression discontinuity (106), time series and multiple baseline comparisons (5). While many of these designs rely on quantitative analysis, qualitative methods can also be used by themselves or in mixed-methods designs that combine qualitative and quantitative methods, to precede, confirm, complement or extend quantitative evaluation of effectiveness (83). Within this growing family of randomized, nonrandomized, and mixed method designs, reasonable

consensus has grown across diverse fields about when certain designs should be used and what sample size requirements and design protocols are necessary to maximize internal validity (42, 87).

Dissemination and implementation research represents a distinct stage, and the designs for this newer field of research are currently less well-established than those for efficacy and effectiveness. Understanding the full range of these designs has impeded the development of dissemination and implementation science and practice. Dissemination and implementation ultimately aim to improve the adoption, appropriate adaptation, delivery, and sustainment of effective interventions by providers, clinics, organizations, communities, and systems of care. In public health, dissemination and implementation research is intimately connected to understanding how the following seven intervention categories can be delivered in and function effectively in different contexts: Programs (e.g. cognitive behavioral therapy), Practices (e.g., “catch-em being good” (84, 96), Principles (e.g., prevention before treatment), Procedures (e.g., screen for depression), Products (e.g., mHealth app for exercise), Pills (e.g., PrEP to prevent HIV infection) (51), and Policies (e.g., limit prescriptions for narcotics). This paper uses the term clinical/preventive intervention to refer to a single or multiple set of these 7 Ps that are intended to improve health for individuals, groups, or populations. Dissemination refers to the targeted distribution of information or materials to a specific public health or clinical audience while implementation involves “the use of strategies to adopt and integrate evidence-based health interventions and change practice patterns within specific settings” (49). Dissemination distributions and implementation strategies may be designed to prevent a disorder or the onset of an adverse health condition, may intercede around the time of this event, may be continuous over a period of time, or may occur afterwards.

Dissemination and implementation research pays explicit (although not exclusive) attention to external validity, in contrast to the main emphasis on internal validity in most randomized efficacy and effectiveness trials (20, 48, 52). Limitations in our understanding of dissemination and implementation have been well-documented (1, 2, 86). But some have called for a moratorium on randomized efficacy trials for evaluating new health interventions until we address the vast disparity between what we know could work under ideal conditions versus what we know about program delivery in practice and in community settings (63).

There is considerable debate as to whether and to what extent designs involving randomized assignment should be used in dissemination and implementation studies (79), as well as the relative contributions of qualitative, quantitative, and mixed methods (83). Some believe there is value in incorporating random assignment designs early in the implementation research process to control for exogenous factors across heterogeneous settings (14, 25, 66). Others are less sanguine about randomized designs in this context and suggest nonrandomized alternatives (71, 102). Debates about research designs for the emerging field of dissemination and implementation are often predicated on conflicting views of dissemination and implementation research and practice, such as whether the evaluation is intended to produce generalizable knowledge, support local quality improvement, or both (27), and the related underlying scientific issue of how much emphasis to place on internal validity compared to external validity (52).

In this paper, we introduce a conceptual view of the traditional translational pipeline that was formulated as a continuum of research originally known as Levy's arrow (68). This traditional translational pipeline is commonly used by the National Institutes of Health (NIH) and other research-focused organizations to move scientific knowledge from basic and other pre-intervention research, to efficacy and effectiveness trials, to a stage that reaches the public (66, 79). By no means does all of dissemination and implementation research follow this traditional translational pipeline, so we mention in the discussion section three different classes of research designs that are of major importance to dissemination and implementation research. We also mention other methodologic issues, as well as community perspectives and partnerships that must be taken into consideration.

This paper is a product of a design workgroup formed in 2013 by NIH in order to address dissemination and implementation research. We established a shared definition of terms that are often used differently across health fields. This required significant compromise, as the same words often have different meanings in different fields. Indeed, the term "design," as used by quantitative or qualitative methodologists and intervention developers, is entirely different. Three terms we use repeatedly are process, output, and outcome. As used here, process refers to activities undertaken by the health system (e.g., frequency of supervision), output refers to observable measures of service delivery provided to the target population, (e.g., the numbers of the eligible population who take medication), and outcome refers only to health, illness, or health-related behaviors of individuals who are the ultimate target of the clinical/preventive intervention. We provide other consensus definitions involving dissemination and implementation and statistical design terms throughout the paper.

### **Where Dissemination and Implementation Fit in the Traditional Translational Pipeline**

An updated version of the National Academy of Medicine (NAM; formerly the Institute of Medicine [IOM]) 2009 perspective on the traditional translational pipeline appears in Figure 1. This top-down translation approach (79) begins with basic and other pre-intervention research at the lower left that can inform the development of novel clinical/preventive interventions. These new interventions are then tested in tightly controlled efficacy trials to assess impact under ideal conditions. Typically, a highly-trained research team would deliver this program to a homogeneous group of subjects with careful monitoring and supervision to ensure high fidelity. Recognizing that efficacy trials can answer only questions of whether a clinical/preventive intervention could work under rigorous conditions, such a program or practice that demonstrates sufficient efficacy would then be followed, in this traditional research pipeline, by an effectiveness trial in the middle of Figure 1, embedded in the community and/or organizational system where such a clinical/preventive intervention would ultimately be delivered. Such effectiveness trials typically have clinicians, other practitioners, or trained individuals from the community deliver the clinical/preventive intervention with ongoing supervision supported by researchers. Also, a more heterogeneous group of study participants are generally included in an effectiveness trial compared to an efficacy trial. These less-controlled conditions allow an effectiveness trial to determine if a clinical/preventive intervention does work in a realistic context.

The final stage of research in this traditional translational pipeline model concerns how to make such a program work within community and/or service settings, the domain of dissemination and implementation research and the last stage of research in the traditional research pipeline. Following this pipeline, a prerequisite for conducting an implementation study would be that the clinical/preventive intervention has already demonstrated effectiveness. Effectiveness of the clinical/preventive intervention in this traditional research pipeline would be considered “settled law” so that proponents of this translational pipeline consider it unnecessary to re-examine effectiveness in the midst of an implementation research design (38, 100, 109). Thus, the traditional translational pipeline model is built around those clinical/preventive interventions that have succeeded in making it through the effectiveness stage.

We now describe the focus of dissemination and implementation research under this traditional research pipeline (see Figure 1 upper right). A tacit assumption of this pipeline is that wide-scale use of evidence-based clinical/preventive interventions generally requires a targeted information dissemination and often a concerted, deliberate strategy for implementation to move to this end of the diffusion, dissemination, and implementation continuum (53, 81, 94). A second assumption is that for a clinical/preventive intervention to have population-level impact, it must not only be an effective program, it must reach a large portion of the population, be delivered with fidelity, and maintained (50). Within the dissemination and implementation research agenda, there are phases of the implementation process itself that have been distinguished. A common exemplar, the “EPIS” conceptual model of the implementation process (1), identifies four phases as represented by the four white boxes within implementation illustrated in Figure 1. The first of these phases, Exploration, refers to whether a service delivery system (e.g., health care, social service, education) or community organization would find a particular clinical/preventive intervention useful for them given their outer context (e.g., service system, federal policy, funding) and inner context (e.g., organizational climate, provider experience). The Preparation phase refers to putting into place the collaborations, policies, funding, supports, and processes needed across the multilevel outer and inner contexts to introduce this new clinical/preventive intervention into this service setting once the adoption decision is made. In this phase, adaptations to the service system, service delivery organizations, and the clinical/preventive intervention itself are considered and prepared. The Implementation (with fidelity) phase refers to the support processes that are developed both within a host delivery system and its affiliates to recruit, train, monitor, and supervise intervention agents to deliver the intervention with adherence and competence and if necessary to adapt systematically to the local context (35). The final phase is Sustainment, referring to how a host delivery system and organizations maintains or extends the supports as well as the clinical/preventive intervention, especially after initial funding has ended. The entire set of structural, organizational, and procedural processes that form the support structure for a clinical/preventive intervention is referred to in this paper as the implementation strategy; this is viewed as distinct from, but generally dependent on, the specific clinical/preventive intervention that is being adopted.

Also shown in this diagram is a contrast between local formative evaluation or quality improvement, and generalizable knowledge, represented by the depth dimension of the

dissemination and implementation box. Local evaluation is generally designed to test and improve the performance of the implementation strategy to deliver the clinical/preventive intervention in that particular setting, with little or limited interest in generalizing their findings to other settings. Implementation studies designed to produce generalizable knowledge contrast in obvious ways with this local evaluation perspective, but systematic approaches to adaptation can provide generalizable knowledge as well. In the traditional translation pipeline, most of the emphasis is on producing generalizable knowledge.

This traditional pipeline does not imply that research always continues to move in one direction as the sequential progression of intervention studies is often cyclical (14). Trial designs may change as well through this pipeline. Efficacy or effectiveness trials nearly always use random assignment, while implementation research often requires tradeoffs between a randomized trial design that can have high internal validity but is difficult to mount, and an observational intervention study that has little experimental control but still may provide valuable information (71).

## Designs for Dissemination and Implementation Strategies

This section examines three broad categories of designs providing within-site, between-site, and within and between-site comparisons of implementation strategies.

### Within-Site Designs

**Post Design of an Implementation Strategy to Adopt an Evidence-Based Clinical/Preventive Intervention in a New Setting**—The simplest—and often most common design, is a post design, which examines health care processes and health care utilization or output after introduction of an implementation strategy focused on the delivery of an evidence-based clinical/preventive intervention in a novel health setting. The emphasis here is on changing health care process and utilization or output rather than health outcomes (i.e., not measures of patient or subject health or illness). As an example, consider the introduction of rapid oral HIV testing in a hospital-based dental clinic, potentially useful based on the sensitivity and specificity of this test. The Centers for Disease Control and Prevention (CDC) has recently proposed that dental clinics could deliver this new technology, and there are real questions about whether such a strategy would be successful. Implementation requires partnering with the dental clinic (Exploration phase), which would need to accept this new mission, and hiring of a full-time HIV counselor to discuss results with patients (Preparation). Here, a key process measure is the rate in which HIV testing is offered to appropriate patients. Two key output measures are the rate at which an HIV test is conducted and the rate of new HIV positives. Blackstock and colleagues' study (8) was successful in getting dental patients to agree to be tested for HIV within the clinic, but it had no comparison group and did not collect pretest rates of HIV testing among all patients. This program did identify some patients who had not previously been tested and were found to be HIV positive.

For implementation of a complex clinical/preventive intervention in a new setting, this post design can be useful in assessing factors predicting program adoption. For example, all of California's county-level child welfare systems were invited to be trained to adopt

Multidimensional Treatment Foster Care, an evidence-based alternative to group care (23). Only a post test was needed to assess adoption and utilization of this program because the sole purveyor of this program clearly knew when and where this program was being used in any California communities (25). Post designs are also useful when new health guidelines or policy changes occur.

**Pre-Post Design of an Implementation Strategy of a Clinical/Preventive Intervention already in Use**—Pre-post studies require information about pre-implementation levels. Some clinical/preventive interventions are already being used in organizations and communities, but do not have the reach into the target population that program objectives require. A pre-post design can assess such changes in reach. In a pre-post implementation design, all sites receive a new implementation strategy for a clinical/preventive intervention that is already being used, with measurement of process or output prior to implementation and after this strategy begins. Effects due to the new implementation strategy are inferred from these pre to post within-site changes. One example of a study using this design is the Veterans Administration's use of the Chronic Care Model for inpatient smoking cessation (61). A primary output measure in this study is the number of prescriptions for smoking cessation. This pre-post design is useful in examining the impact of a complex implementation strategy within a single organization or across multiple sites that are representative of a population (e.g., federally qualified health centers).

Pre-post designs are also useful in assessing the adoption by health care systems of a guideline, black box warning, or other directive. For example, the comparison of management strategies used for inpatient cellulitis and cutaneous abscess were compared for all patients with these discharge diagnoses in the year prior to and after guidelines were published (8). The effects on prescribing antidepressants to depressed youth due to a black box warning was evaluated by comparing prescription rates (46) and adverse event reports (47) prior to and after the warning.

A variant of the pre-post design involves a multiple baseline time-series design. Biglan and colleagues examined the prevalence of tobacco sales across time to young people without checking birthdays after an outlet store reward and reminder system was implemented (7).

### Between-Site Designs

In contrast to previous designs that examined changes over time within one or more sites exposed to the same dissemination or implementation strategy, these designs compare processes and output among sites having different exposures.

#### **New Implementation Strategy Versus Usual Practice Implementation Designs**

—In new versus Implementation As Usual (IAU) designs, some sites receive an innovative implementation strategy while others maintain their usual condition. Process and output measures can then be compared between the two types of sites over the same period of time. It is possible to use such a design even when one site introduces a new dissemination or implementation approach. As a policy dissemination example, Hahn and colleagues compared county-level smoking rates before and after the enactment of a smoke-free law in one county, comparing this county's response to 30 comparison counties with similar

demographics (54). The effects of the law are evaluated using a regression displacement design (111) that uses multiple time points to examine changes in trajectories for those receiving this policy intervention compared to those that do not.

One illustration where self-selection is important comes from an educational outreach strategy to affect prescribing practices of physicians in managing the bacterium *Helicobacter pylori* (55). Part of this study involved a randomization of practices; those assigned to the active intervention arm were to receive educational outreach and auditing. However, only one half of these practices accepted the educational outreach and only 8% permitted the audit, thus severely limiting the value of the randomized trial. It is still possible to compare prescribing differences among those practices that did receive educational outreach and those that did not, and this was the primary purpose of the paper (55). Interpretation of such differences in this now observational study may be challenging as there may be confounding by selection factors that distinguish those practices that were willing and those that were not willing to receive educational outreach. Principles of diffusion of innovation could help understand why some organizations adopt an implementation strategy while others do not (97). Propensity scores, based on site-level covariates, are useful in controlling for some degree of assignment bias (98, 104).

In a randomized New versus IAU trial, the goal is to determine whether the new implementation strategy produces better or more efficient process and outputs, (e.g., improved reach or penetration of the innovation, or improved utilization of a health care standard or innovation) compared to what now exists (41). Here the sites (practices, communities, organizations) are assigned randomly to the new active implementation or usual practice condition. As random assignment is at the group rather than individual level, this forms a cluster-randomized design (75, 95). Process and output measures used as the primary endpoints are measured for all eligible patients or subjects in both conditions and aggregated to the level of the randomized unit. Such a randomized implementation trial tests whether the new implementation strategy increases utilization of a health care innovation.

One example of a successful evaluation of an implementation strategy is the PROSPER study, which randomly assigned 28 communities to receive a supportive strategy for implementing a combination of evidence-based school and family preventive interventions for youth substance abuse, or to usual practice in the community. With this design, the investigators were able to evaluate initial adoption as well as sustainment and fidelity across multiple cohorts (103).

Instead of larger health organizational sites being randomized, smaller units can sometimes be randomized within each organization. Randomization could occur at the level of the ward, team, clinician, or even at the level of patients or subjects within each site, again assessing health care service utilization as the primary outcome. Such a design uses site as a blocking factor in contrast to the design described above. For example, if similar teams exist and work relatively independently within each organization, an efficient design is to randomly assign teams within each organization into blocks so that a precise comparison can be made between implementation strategies within each organization (12). With complex, multilevel implementation strategies involving the adoption of clinic-level practices, there is



a potential for contamination if two implementation conditions are tested in the same site. For example, if one were to test the introduction of system-level policies for practitioners to increase hand washing at the bedside, a design that randomized small subunits within the system would not be able to test a fully implemented systems approach. A useful rule of thumb is to randomize at the “level of implementation,” that is, at the level where the full impact of the strategy is designed to occur (12). In a recent experiment that tested a hand washing implementation strategy, there was feedback at ward-level meetings (43). Thus, a ward-level randomization was appropriate for this trial.

A few implementation strategies can be tested with randomization even down to the individual patient level. These are strategies where a) leakage of the implementation to other patients is likely to be minimal, and b) the implementation’s impact is not likely to be attenuated due to its exposure on only a portion of patients. An example is the use of automated systems to screen and/or refer patients. Minimal leakage is likely to occur with strategies that involve automated messages to patients, so individual-level randomized designs are appropriate. A recent review of system, practitioner, and individual patient-level randomized trials to evaluate level-specific lipid management approaches that all used a health information technology component concluded that systems and individual patient implementations showed better management (4).

One type of new versus usual practice randomized design for implementation is built around the theme of encouraging system-level health behavior through incentives for desired behavior or penalties for undesired behavior. In these randomized encouragement implementation designs, one strategy receives more attention or incentives than the other. The effect of providing incentives or supports for succeeding to implement or penalties for failing to implement can be evaluated with such a design. An example involves the pay for performance (P4P) strategy to increase therapist adherence to a protocol for adolescent substance abusers (45), as therapists receive financial compensation if they achieve a specified level of competence and have the adolescent complete treatment sessions.

**Head-to-Head Randomized Implementation Trial Design**—A head-to-head randomized implementation trial is a comparative effectiveness implementation trial that tests which of two active, qualitatively different implementation strategies is more successful in implementing a clinical/preventive intervention. Here the same clinical/preventive intervention is used for both arms of this trial, and health or service system units are assigned randomly to one of the two different implementation strategies as shown in Figure 2. Both implementation strategies are manualized and carried out with equivalent attention to fidelity. The two implementation strategies are compared on the quality, quantity, or speed of implementing the clinical/preventive intervention (11).

An example of the use of such a head-to-head randomized implementation trial is the CAL-OH trial that compares two alternative strategies of a county-level implementation of Multidimensional Treatment Foster Care (MTFC) (23, 26), an evidence-based program for foster children and their families that is conducted in the child welfare, juvenile justice, and mental health public service systems in California and Ohio. A total of 51 counties were randomized to one of two implementation conditions: an individual county implementation

strategy (IND), or a community development team (CDT) involving multiple counties in a learning collaborative (11). A diagram of the design of this trial is shown in Figure 3. Counties were assigned randomly to a cohort, which governed when they would start the implementation process, as well as to an implementation strategy condition. This type of roll-out design, where counties' start times were staggered, was chosen because it balanced the demand for training the counties with the supply of training resources available by the purveyor. Counties were matched across a wide range of baseline characteristics so that cohorts formed equivalent blocks. A Stages of Implementation Completion (SIC) measure (24) was developed and used to evaluate the implementation process including the quality of preparedness and training to deliver MTFC, the speed that milestones were achieved, and the quantity of eligible families served (11). Head-to-head testing involved analyzing how combinations of the SIC items' distributions varied by implementation condition (11).

Another type of head-to-head trial design involves two implementation strategies that target different outcomes, with no site receiving both. This design allows each site to serve as both an active intervention and a control because the two implementation targets focus on different patient populations. An example is the simultaneous testing of two clinical pathway strategies in emergency departments for pediatric asthma and pediatric gastroenteritis (59). Sixteen emergency departments are to be randomly assigned to one of these strategies, and key clinical output measures are assessed for asthma and gastroenteritis.

Dosage trials, which assign units to varying intensities of an intervention, are common in efficacy and effectiveness studies, but they can also be used with varying implementation intensity. An example is a trial focused on in-service training and supervision of first-grade teachers in the Good Behavior Game (GBG) to manage classroom behavior. For this trial, first-grade students within schools were assigned randomly to classrooms, and classrooms/teachers were assigned randomly to no training, a low intensity GBG training and supervision, or a more intensive level of supervision with a coach (88, 89).

**Designs for a Suite of Evidence-Based Clinical/Preventive Interventions—**Up to now, this typology focuses on a single evidence-based clinical/preventive intervention. This one-choice option does not allow communities or organizations to select programs that match their needs, values, and resources. A decision support system to select evidence-based programs is, in fact, an implementation strategy, and such a support system can also be tested for impact using a well-crafted design.

An example of a randomized implementation trial of such a decision support system for prevention of youth substance use and violence is the Community Youth Development Study (CYDS). This randomized trial of 24 communities (18, 76) tested the Communities that Care (CTC) (57) comprehensive community support system against community control condition, which only received information about their community's risk and protective factor profile but no technological help in determining which programs would be successful. This program measured implementation process milestones and benchmarks and compared both outputs, including the number of evidence-based prevention programs adopted by these communities and drug use and violence outcomes at the community level (19, 56, 73, 82). This study

found that the CTC decision support system led to greater adoption of evidence-based programs and prevented youth substance use and violence.

**Factorial Designs for Implementation**—Factorial designs for implementation investigate the combination of two or more implementation strategies at a time. Each experimental factor has two or more levels (e.g., presence or absence; low, medium, high intensity). A 2×2 factorial implementation design assigns units randomly to one of the four conditions and provides estimates of each factor by itself and their interaction. An example of a 3×3 design is the evaluation of three alternative screening tools for alcohol abuse in emergency departments, used in combination with three types of advice by an alcohol health worker (minimal intervention, brief advice, brief advice plus counseling). Individual emergency departments are randomized to a single screening tool and level of advice (33).

With incomplete factorial designs, one or more arms of a complete factorial are excluded from the study. For example, in a design involving the presence or absence of two implementation strategies, it may be viewed as unethical to withhold both strategies; thus, units may be assigned to either implementation or both. Alternatively, it may be too complex or unmanageable to conduct both implementation strategies in the same unit, therefore excluding the combined strategy.

We can consider testing a large number of components that go into an implementation strategy using Multiphase Optimization Strategy Implementation Trials (MOST) (29, 32, 69, 85, 91, 112). This approach recognizes that many choices can be made in an implementation strategy. For example, a comprehensive implementation strategy often requires components that involve system leadership, the clinic level, the clinician level, as well as key processes: planning, educating, financing, restructuring, managing quality, and/or policy (90). Many of these components are thought to be necessary for an implementation strategy to work properly. With diverse approaches in the ways that these implementation components can be specified (92) and connected (105), as well as variations in strength, there is an exponential explosion of possible implementation strategies that can be developed. Testing all combinations in a single design is not feasible, but MOST can be used to identify and test an optimized intervention. MOST has three phases. The first phase, preparation, involves selection and pilot testing components with a clear optimization criterion (e.g., most effective components subject to a maximum cost). In the second phase, optimization, a fully powered randomized experiment is conducted to assess the effectiveness of each intervention component. The number of distinct implementation strategies is minimized using a balanced, fractional factorial experiment. Fractional designs can make examination of multiple components feasible, even when cluster randomization is necessary (30, 32, 37). The set of components that best meets the optimization criterion is identified based on the trial's results. In the third phase, evaluation, a standard randomized implementation trial is conducted comparing the optimized intervention against an appropriate comparison condition.

An example of a MOST implementation trial involved determining which of a set of three components improved the fidelity of teacher delivery of Healthwise (21, 22) for use in South Africa. The three components were school climate, teacher training, and teacher supervision.

These three components were tested in a factorial experiment, assigning 56 schools to one of eight experimental conditions (22).

The Sequential Multiple Assignment Randomized Implementation Trial (SMART) implementation design is a special case of the factorial experiment (32, 67, 78) involving multistage randomizations where the site-level implementation process can be modified if unsuccessful. Such an adaptive approach to enhancing implementation can make the best allocation of available resources (31) and change its approach if a strategy is failing. For example, the Replicating Effective Programs (REP) strategy was developed to promote proper implementation of evidence-based health care interventions in community settings (80), but one study found that fewer than half of the sites sustained use of evidence-based interventions (65). In a subsequent study, Kilbourne et al. (2014) are using SMART to examine an adaptive version of REP (64). Initially, 80 community-based outpatient clinics are provided with REP. Clinics that do not respond are randomized to receive additional support from an external facilitator only, or both an external and internal facilitator. Clinics that are randomized to receive an external facilitator only and are still unresponsive will be re-randomized to continue with only an external facilitator or to add an internal facilitator.

In SMART designs, the adaptation probability within each site does not vary. Cheung et al. (2015) propose a SMART design with adaptive randomization (SMART-AR) in which the randomization probabilities of the strategies are updated based on outcomes in other sites, so as to improve the expected outcome (28). Such adaptations are potentially infinite if the interventions are complex and the intended populations and settings highly varied.

**Doubly Randomized, Two-Level Nested Designs for Testing Two Nested Implementation Factors**—In a doubly randomized two-level nested or split-plot implementation design, there are two experimental implementation factors directed towards two distinct hierarchical levels—for example, the practitioner and organization. A doubly randomized nested design can test whether additive or synergistic effects exist across the two levels. An example of this design is a trial for smoking cessation implementation, testing whether direct patient reimbursement for medication and/or physician group training influence the use of these medications. Patients are nested within physicians, and all four combinations are tested (107).

### **Within- and Between-Site Comparison Designs**

Within- and between-site comparisons can be made with crossover designs where sites begin in one implementation condition and move to another. We use the generic term of a roll-out randomized implementation trial to refer to the broad class of designs where the timing of the start of an implementation strategy is randomly assigned. In the simplest roll-out design, all units start in a usual practice setting, then they cross over at randomly determined time intervals, and all eventually receive a new implementation strategy. Thus, at each time interval there is a between-units comparison of those units that are assigned to the new implementation strategy and those that remain as a usual practice setting. In addition, a within-unit comparison can be made as units change from a usual practice setting to a new implementation strategy at some randomly assigned time.

Figure 4(a) shows this simplest type of roll-out design where all units are identified at the beginning of the study; all begin in usual practice, designated by 0 in this figure, and are randomly assigned to a cohort, i.e. when they are to receive the implementation strategy (time 1, 2, 3, or 4), designated by X\*. After this startup period, the implementation strategy continues across the remaining time periods for those units that have crossed over to the new implementation strategy. We have designated this by X. Measures are taken across all cohorts at all time periods. This type of roll-out design has been used for about two decades in effectiveness trials, where it is known as a dynamic wait-listed design (17) or a stepped-wedge design (10). Communities and organizations often are willing to accept this type of roll-out design over a traditional design when there are obvious or perceived advantages in receiving the new implementation strategy, or when it is unethical to withhold a new implementation strategy throughout the study (17).

Other roll-out designs can be used for implementation (111). In Figure 4(b), all sites start in an IAU and then at a random time are assigned to start one of two new implementation strategies, labeled X\* and Y\* in Figure 4(b). They continue in this same condition until the end of the study. A design such as this is being used in an ongoing trial in the juvenile justice system conducted by the National Institute on Drug Abuse (NIDA) and colleagues (6). We call this a head-to-head roll-out trial to distinguish it from stepped-wedge designs discussed above.

In addition, other types of roll-out designs could be used, but to our knowledge have not yet been used in implementation. A pairwise enrollment roll-out design (111) differs from the two we previously discussed and is similar to the original design for a roll-out randomized effectiveness trial of an HIV community-based intervention. In their original design (62), a pair of communities were randomized to receive the intervention in the first or second year; assessments were made for both communities in the first year to evaluate impact. This pairwise randomization would be repeated in following years so that over time a true randomized trial with sufficient numbers of units could be conducted. This type of pairwise enrollment roll-out design shown in Figure 4(c) could also work for implementation trials, thereby eliminating the need to enumerate all sites at the beginning of the study.

## Discussion

In this paper, we have provided an extensive but admittedly incomplete compendium of designs that are or could be used in dissemination and implementation research under the general classification of the traditional translational pipeline. These are suitable for many clinical/preventive interventions that are judged to have successfully transcended the scale of evidence through effectiveness. This pipeline is useful in implementing a predetermined clinical/preventive intervention, or ones from a suite of evidence-based programs. Thus, these clinical/preventive interventions should be standardized and stable, subject to limited adaptation rather than allowed to change drastically. We presented designs for this pipeline involving within-site comparisons alone, between-site comparisons, and within- and between-site comparisons with roll-out designs. A wide range of factorial designs can also be used to evaluate multiple implementation components. Many of the designs we described involved randomization, which can often strengthen inferences, but there are many situations

where randomization is not possible or required, nor even advisable. For example, there are times when the policies are disseminated or implemented by law (93) or other nonrandom process. Likewise, dissemination or implementation may involve one single community or organization, with a research focus on understanding the internal diffusion process (110) involving network connections inside this system (108).

There are many excellent dissemination and implementation designs that address issues beyond those relevant to the traditional translational pipeline, and our focus on pipeline designs should in no way be interpreted as reducing the importance of these other designs. We note three broad types not discussed: those where effectiveness of the clinical preventive intervention is to be evaluated alongside questions involving implementation, which include hybrid designs (36) and continuously evolving interventions (72), those designs that address quality improvement for local knowledge (27), and designs involving simulation or synthetic experiments in dissemination and implementation (44). These designs will be presented elsewhere.

We did not provide details on statistical power in this review, despite its obvious importance. A few general comments on power should be made. Dissemination and implementation research, because it often tests system-level strategies, generally requires multilevel data with sizeable numbers of group level (e.g., practices or organizations) as well as individuals, as the statistical power of such designs is most strongly influenced by the number units at the highest level (75, 77). Some general approaches to increasing statistical power include blocking and matching within the design itself and analytical adjustment with covariates at the unit of randomization to reduce imbalance. Triangulating the findings of quantitative analyses with qualitative data in mixed method designs is another approach to addressing limited statistical power (83).

Neither have we discussed statistical analyses or causal inference related to specific designs. Again, we point out a few issues. First, it is common in implementation trials for many of the units assigned to an implementation strategy to end up not adopting the intervention. In such cases, there are analytic ways to account for self-selection factors related to adoption. One analytical approach to accounting for incomplete adoption is to use Complier Average Causal Effect (commonly known as CACE) modeling (3, 9, 60). Secondly, a primary outcome for most of these designs can be constructed as a composite process and output score, using dimensions of speed, quality and quantity (11, 25, 101).

In terms of causality, nonrandomized designs generally provide less confidence that the numerical comparisons are due to the differences in implementation conditions. For example, the effects of examining a new implementation in a post-test only design can be confounded with changes in multiple policies and other external factors, whereas a randomized new versus IAU maintains some protection. Such designs cannot completely rule out all unmeasured external effects the way randomized trials can. There are also some subtle issues that arise from group-randomized implementation trials (16). First is a conceptual problem of what is meant by causality at a group level. The standard assumptions put forward in some causal inference paradigms involving experimental trials are generally not valid for dissemination and implementation designs. Specifically, the assumption that a

subunit's own output or outcomes are not affected by any other unit's implementation assignment (what is called SUTVA (99)) is invalid, as interactions and synergistic effects underlie many of our implementation strategies. In addition, implementation strategies are inherently systemic rather than linear, and to date there is no fully-developed causal inference approach comparable to that of single-person randomized trials that addresses this cyclic nature of implementation. Despite these theoretical concerns, randomized dissemination and implementation trials can provide valuable information for policy and practice, as well as generalizable knowledge.

While we have provided some illustrations of the use of many of the designs discussed above, this paper does not provide recommendations regarding the appropriateness of one design over another. The choice of design is a complex process, with major input regarding the research questions; the state of existing knowledge; the intention to obtain generalizable or local knowledge; the community, organizational, and funder values, expectations, and resources; and the available opportunities to conduct such research (71). All of these factors are beyond the scope of this paper. Nevertheless, the designs listed here can be considered as options for researchers, funders, and organizational and community partners alike.

We close by reinforcing the message that researchers and evaluators represent only one sector of people that make design decisions in implementation. More than efficacy and effectiveness studies, dissemination and implementation studies involve major changes in organizations and communities, and as such, community leaders, organizational leaders, and policy makers have far more at stake than the evaluators. The most attractive scientific design on paper will not happen without the endorsement and agreement of the communities and organizations where this takes place.

## Acknowledgments

We are grateful for support from the National Institute on Drug Abuse (NIDA) (P30DA027828, C Hendricks Brown PI), the National Institute of Mental Health (NIMH) (R01MH076158, Patricia Chamberlain, PI; R01MH072961, Gregory Aarons, PI). This paper grew out of a NIDA, NIMH, and NCI sponsored workgroup, as part of the 6<sup>th</sup> Annual NIH Meeting on Advancing the Science of Dissemination & Implementation Research: Focus on Study Designs. Earlier versions of this paper were presented as a webinar, and at the 7<sup>th</sup> Annual NIH Meeting on Advancing the Science of Dissemination & Implementation Research. NIH staff received no support from extramural grants for their involvement. We thank the reviewers for many helpful comments. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

## LITERATURE CITED

1. Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health*. 2011; 38:4–23. [PubMed: 21197565]
2. Aarons GA, Sommerfeld DH, Walrath-Greene CM. Evidence-based practice implementation: the impact of public versus private sector organization type on organizational support, provider attitudes, and adoption of evidence-based practice. *Implementation science: IS*. 2009; 4:83–5908-4-83.
3. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*. 1996; 91:444–55.

4. Aspry KE, Furman R, Karalis DG, Jacobson TA, Zhang AM, et al. Effect of health information technology interventions on lipid management in clinical practice: A systematic review of randomized controlled trials. *Journal of Clinical Lipidology*. 2013; 7:546–60. [PubMed: 24314354]
5. Baer DM, Wolf MM, Risley TR. Some current dimensions of applied behavior analysis. *Journal of applied behavior analysis*. 1968; 1:91–97. [PubMed: 16795165]
6. Belenko S, Wiley T, Knight D, Dennis M, Wasserman G, Taxman F. A new behavioral health services cascade framework for measuring unmet addiction health services needs and adolescent offenders: conceptual and measurement challenges. *Addiction Science & Clinical Practice*. 2015; 10:A4.
7. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. *Prevention Science*. 2000; 1:31–49. [PubMed: 11507793]
8. Blackstock OJ, King JR, Mason RD, Lee CC, Mannheimer SB. Evaluation of a rapid HIV testing initiative in an urban, hospital-based dental clinic. *AIDS Patient Care and STDs*. 2010; 24:781–85. [PubMed: 21091231]
9. Bloom HS. Accounting for no-shows in experimental evaluation designs. *Evaluation Review*. 1984; 8:225–46.
10. Brown CA, Lilford RJ. The stepped wedge trial design: A systematic review. *BMC Medical Research Methodology*. 2006; 6:54. [PubMed: 17092344]
11. Brown CH, Chamberlain P, Saldana L, Padgett C, Wang W, Cruden G. Evaluation of two implementation strategies in fifty-one child county public service systems in two states: Results of a cluster randomized head-to-head implementation trial. *Implementation Science*. 2014; 9:134. [PubMed: 25312005]
12. Brown CH, Liao J. Principles for designing randomized preventive trials in mental health: An emerging developmental epidemiology paradigm. *American Journal of Community Psychology*. 1999; 27:673–710. [PubMed: 10676544]
13. Brown CH, Mohr DC, Gallo CG, Mader C, Palinkas L, et al. A Computational Future for Preventing HIV in Minority Communities: How Advanced Technology Can Improve Implementation of Effective Programs. *JAIDS Journal of Acquired Immune Deficiency Syndromes*. 2013; 63:S72–S84. [PubMed: 23673892]
14. Brown CH, Ten Have TR, Jo B, Dagne G, Wyman PA, et al. Adaptive designs for randomized trials in public health. *Annual Review of Public Health*. 2009; 30:1–25.
15. Brown CH, Wang W, Kellam SG, Muthén BO, Petras H, et al. Methods for testing theory and evaluating impact in randomized field trials: Intent-to-treat analyses for integrating the perspectives of person, place, and time. *Drug and Alcohol Dependence*. 2008; 95:S74–S104. [PubMed: 18215473]
16. Brown CH, Wyman PA, Brinales JM, Gibbons RD. The role of randomized trials in testing interventions for the prevention of youth suicide. *International Review of Psychiatry*. 2007; 19:617–31. [PubMed: 18092240]
17. Brown CH, Wyman PA, Guo J, Peña J. Dynamic wait-listed designs for randomized trials: new designs for prevention of youth suicide. *Clinical Trials*. 2006; 3:259–71. [PubMed: 16895043]
18. Brown E, Hawkins J, Arthur M, Briney J, Abbott R. Effects of Communities That Care on Prevention Services Systems: Findings From the Community Youth Development Study at 1.5 Years. *Prevention Science*. 2007; 8:180–91. [PubMed: 17602298]
19. Brown E, Hawkins JD, Rhew I, Shapiro V, Abbott R, et al. Prevention System Mediation of Communities That Care Effects on Youth Outcomes. *Prevention Science*. 2014; 15:623–32. [PubMed: 23828448]
20. Brownson RC, Roux AVD, Swartz K. Commentary: Generating Rigorous Evidence for Public Health: The Need for New Thinking to Improve Research and Practice. *Annual Review of Public Health*. 2014; 35:1–7.
21. Caldwell LL, Patrick ME, Smith EA, Palen L-A, Wegner L. Influencing adolescent leisure motivation: Intervention effects of HealthWise South Africa. *Journal of Leisure Research*. 2010; 42:203. [PubMed: 25429164]



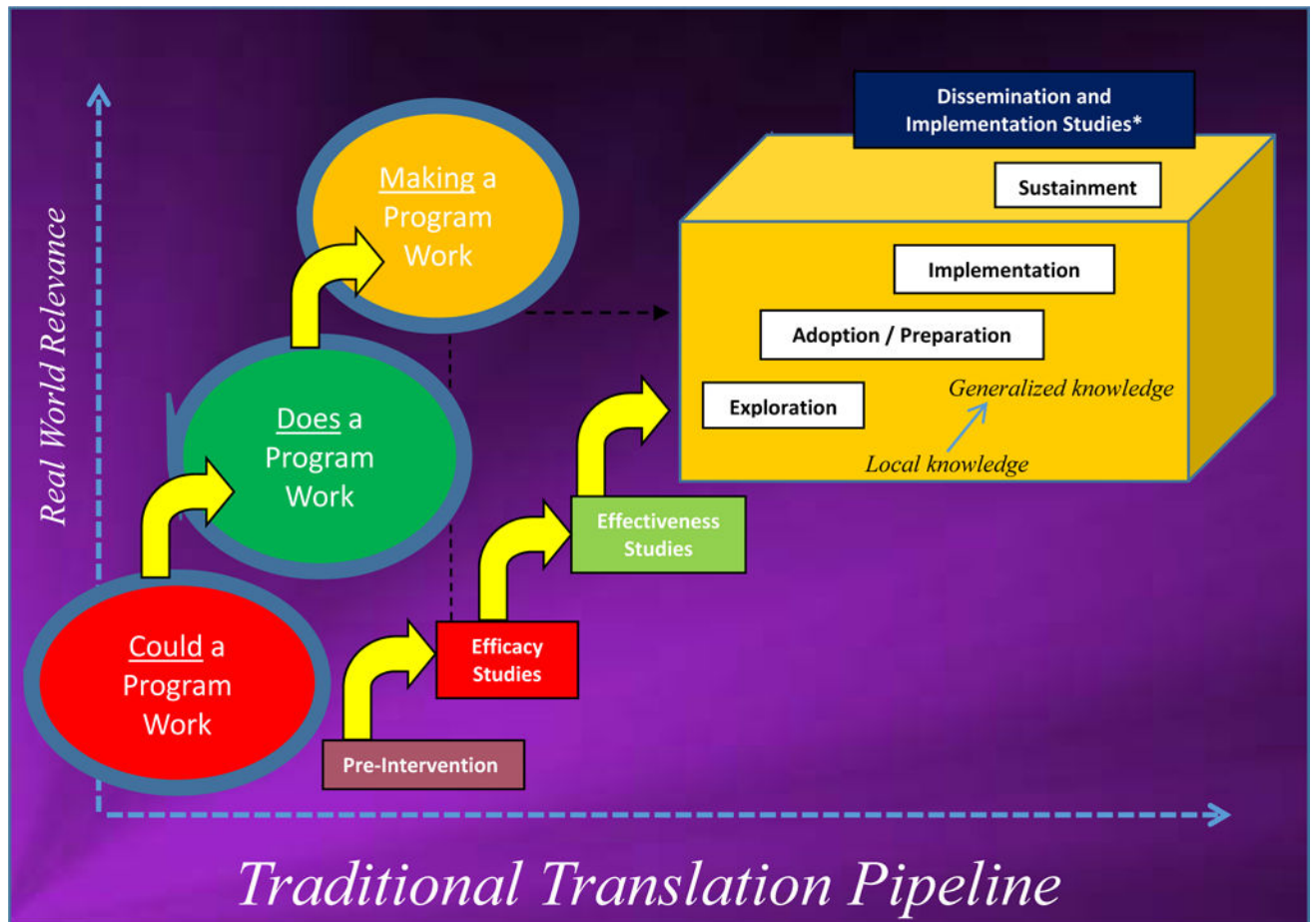
22. Caldwell LL, Smith EA, Collins LM, Graham JW, Lai M, et al. Translational research in South Africa: Evaluating implementation quality using a factorial design. Presented at Child & youth care forum. 2012
23. Chamberlain, P. Treating chronic juvenile offenders: Advances made through the Oregon multidimensional treatment foster care model. American Psychological Association; Washington, DC: 2003.
24. Chamberlain P, Brown CH, Saldana L. Observational measure of implementation progress in community based settings: The Stages of Implementation Completion (SIC). *Implementation Science*. 2011; 6:116. [PubMed: 21974914]
25. Chamberlain P, Brown CH, Saldana L, Reid J, Wang W, et al. Engaging and recruiting counties in an experiment on implementing evidence-based practice in California. *Admin Policy Ment Hlth Ment Hlth Res*. 2008; 35:250–60.
26. Chamberlain P, Price J, Leve LD, Laurent H, Landsverk JA, Reid JB. Prevention of behavior problems for children in foster care: outcomes and mediation effects. *Prevention Science*. 2008; 9:17–27. [PubMed: 18185995]
27. Cheung K, Duan N. Design of Implementation Studies for Quality Improvement Programs: An Effectiveness-Cost-Effectiveness Framework. *Am J Public Health*. 2013:e1–e8.
28. Cheung YK, Chakraborty B, Davidson KW. Sequential multiple assignment randomized trial (SMART) with adaptive randomization for quality improvement in depression treatment program. *Biometrics*. 2014 ePub ahead of print.
29. Collins LM, Baker TB, Mermelstein RJ, Piper ME, Jorenby DE, et al. The Multiphase Optimization Strategy for Engineering Effective Tobacco Use Interventions. *Annals of Behavioral Medicine : A Publication of the Society of Behavioral Medicine*. 2011; 41:208–26. [PubMed: 21132416]
30. Collins LM, Dziak JJ, Li R. Design of experiments with multiple independent variables: a resource management perspective on complete and reduced factorial designs. *Psychological methods*. 2009; 14:202. [PubMed: 19719358]
31. Collins LM, Murphy SA, Bierman KL. A conceptual framework for adaptive preventive interventions. *Prevention Science*. 2004; 5:185–96. [PubMed: 15470938]
32. Collins LM, Nahum-Shani I, Almirall D. Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART). *Clinical trials (London, England)*. 2014; 11:426–34.
33. Coulton S, Perryman K, Bland M, Cassidy P, Crawford M, et al. Screening and brief interventions for hazardous alcohol use in accident and emergency departments: a randomised controlled trial protocol. *BMC health services research*. 2009; 9 114-6963-9-114.
34. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. *British Medical Journal (Clinical research ed)*. 2008; 337:a1655.
35. Cross W, West J, Wyman PA, Schmeelk-Cone K, Xia Y, et al. Observational Measures of Implementer Fidelity for a School-based Prevention Intervention: Development, Reliability, and Validity. *Prevention Science*. 2015; 16:122–32. [PubMed: 24736951]
36. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Medical care*. 2012; 50:217–26. [PubMed: 22310560]
37. Dziak JJ, Nahum-Shani I, Collins LM. Multilevel factorial experiments for developing behavioral interventions: power, sample size, and resource considerations. *Psychological methods*. 2012; 17:153. [PubMed: 22309956]
38. Elliott DS, Mihalic S. Issues in disseminating and replicating effective prevention programs. *Prevention Science*. 2004; 5:47–53. [PubMed: 15058912]
39. Fisher, RAS. The design of experiments. Edinburgh: Oliver and Boyd; 1935.
40. Flay BR. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine*. 1986; 15:451–74. [PubMed: 3534875]

41. Folks B, LeBlanc WG, Staton EW, Pace WD. Reconsidering low-dose aspirin therapy for cardiovascular disease: A study protocol for physician and patient behavioral change. *Implementation Science*. 2011; 6
42. Friedman, LM., Furberg, C., DeMets, DL. *Fundamentals of clinical trials*. St. Louis: Mosby-Year Book; 1996. p. xviii, 361
43. Fuller C, Michie S, Savage J, McAteer J, Besser S, et al. The Feedback Intervention Trial (FIT)—Improving Hand-Hygiene Compliance in UK Healthcare Workers: A Stepped Wedge Cluster Randomised Controlled Trial. *PloS one*. 2012; 7:e41617. [PubMed: 23110040]
44. Gallo C, Villamar JA, Cruden G, Beck EC, Mustanski B, et al. Systems-level methodology for conducting implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*. In Press.
45. Garner BR, Godley SH, Dennis ML, Hunter BD, Bair CML, Godley MD. Using pay for performance to improve treatment implementation for adolescent substance use disorders: results from a cluster randomized trial. *Archives of Pediatrics & Adolescent Medicine*. 2012; 166:938–44. [PubMed: 22893231]
46. Gibbons RD, Brown CH, Hur K, Marcus SM, Bhaumik DK, et al. Early evidence on the effects of regulators' suicidality warnings on SSRI prescriptions and suicide in children and adolescents. *American Journal of Psychiatry*. 2007; 164:1356–63. [PubMed: 17728420]
47. Gibbons RD, Segawa E, Karabatsos G, Amatya AK, Bhaumik DK, et al. Mixed-effects Poisson regression analysis of adverse event reports: the relationship between antidepressants and suicide. *Statistics in Medicine*. 2008; 27:1814–33. [PubMed: 18404622]
48. Glasgow RE, Magid DJ, Beck A, Ritzwoller D, Estabrooks PA. Practical clinical trials for translating research to practice: design and measurement recommendations. *Medical care*. 2005; 43:551–57. [PubMed: 15908849]
49. Glasgow RE, Vinson C, Chambers D, Khoury MJ, Kaplan RM, Hunter C. National Institutes of Health Approaches to Dissemination and Implementation Science: Current and Future Directions. *American Journal of Public Health*. 2012; 102:1274–81. [PubMed: 22594758]
50. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *American Journal of Public Health*. 1999; 89:1322–27. [PubMed: 10474547]
51. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *New England Journal of Medicine*. 2010; 363:2587–99. [PubMed: 21091279]
52. Green LW, Glasgow RE. Evaluating the Relevance, Generalization, and Applicability of Research: Issues in External Validation and Translation Methodology. *Evaluation & the health professions*. 2006; 29:126–53. [PubMed: 16510882]
53. Greenhalgh T, Robert G, MacFarlane F, Bate P, Kyriakidou O. Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations. *The Milbank quarterly*. 2004; 82:581–629. [PubMed: 15595944]
54. Hahn EJ, Rayens MK, Butler KM, Zhang M, Durbin E, Steinke D. Smoke-free laws and adult smoking prevalence. *Preventive Medicine*. 2008; 47:206–09. [PubMed: 18519154]
55. Hall L, Eccles M, Barton R, Steen N, Campbell M. Is untargeted outreach visiting in primary care effective? A pragmatic randomized controlled trial. *Journal of public health medicine*. 2001; 23:109–13. [PubMed: 11450926]
56. Hawkins DJ, Oesterle S, Brown EC, Abbott RD, Catalano RF. Youth problem behaviors 8 years after implementing the communities that care prevention system: A community-randomized trial. *JAMA Pediatrics*. 2014; 168:122–29. [PubMed: 24322060]
57. Hawkins, JD., Catalano, RF. *Investing in your community's youth: An introduction to the Communities that Care system*. South Deerfield, MA: Channing Bete Company; 2002.
58. Hill, ABS. *Principles of medical statistics*. London: Lancet; 1961.
59. Jabbour M, Curran J, Scott SD, Guttman A, Rotter T, et al. Best strategies to implement clinical pathways in an emergency department setting: study protocol for a cluster randomized controlled trial. *Implementation science : IS*. 2013; 8 55-5908-8-55.

60. Jo B, Asparouhov T, Muthen BO, Jalongo NS, Brown CH. Cluster randomized trials with treatment noncompliance. *Psychological Methods*. 2008; 13:1–18. [PubMed: 18331150]
61. Katz D, Vander Weg M, Fu S, Prochazka A, Grant K, et al. A before-after implementation trial of smoking cessation guidelines in hospitalized veterans. *Implementation Science*. 2009; 4:58. [PubMed: 19744339]
62. Kegeles SM, Hays RB, Pollack LM, Coates TJ. Mobilizing young gay and bisexual men for HIV prevention: a two-community study. *AIDS*. 1999; 13:1753–62. [PubMed: 10509578]
63. Kessler R, Glasgow RE. A proposal to speed translation of healthcare research into practice: dramatic change is needed. *American Journal of Preventive Medicine*. 2011; 40:637–44. [PubMed: 21565657]
64. Kilbourne AM, Almirall D, Eisenberg D, Waxmonsky J, Goodrich DE, et al. Protocol: Adaptive Implementation of Effective Programs Trial (ADEPT): cluster randomized SMART trial comparing a standard versus enhanced implementation strategy to improve outcomes of a mood disorders program. *Implementation Science: IS*. 2014; 9doi: 10.1186/s13012-014-0132-x
65. Kilbourne AM, Neumann MS, Pincus HA, Bauer MS, Stall R. Implementing evidence-based interventions in health care: application of the replicating effective programs framework. *Implementation science: IS*. 2007; 2 42-5908-2-42.
66. Landsverk, J., Brown, CH., Chamberlain, P., Palinkas, L., Ogihara, M., et al. Design and analysis in dissemination and implementation research. In: Brownson, RC, Colditz, GA., Proctor, EK., editors. *Dissemination and Implementation Research in Health: Translating Science to Practice*. London: Oxford University Press; 2012. p. 225-60.
67. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy S. A “SMART” design for building individualized treatment sequences. *Annual review of clinical psychology*. 2012; 8:21–48.
68. Levy RI. The National Heart, Lung, and Blood Institute: Overview 1980: The Director’s Report to the NHLBI Advisory Council. *Circulation*. 1982; 65:217–25. [PubMed: 7032745]
69. McClure JB, Derry H, Riggs KR, Westbrook EW, St John J, et al. Questions About Quitting (Q(2)): Design and Methods of a Multiphase Optimization Strategy (MOST) Randomized Screening Experiment for an Online, Motivational Smoking Cessation Intervention. *Contemporary clinical trials*. 2012; 33:1094–102. [PubMed: 22771577]
70. Medical Research C. A framework for the development and evaluation of RCTs for complex interventions to improve health. MRC. 2000
71. Mercer SL, DeVinney BJ, Fine LJ, Green LW, Dougherty D. Study designs for effectiveness and translation research: identifying trade-offs. *American Journal of Preventive Medicine*. 2007; 33:139–54. [PubMed: 17673103]
72. Mohr DC, Schueller SM, Riley WT, Brown CH, Cuijpers P, et al. Trials of intervention principles: Evaluation methods for evolving behavioral intervention technologies. *Journal of Medical Internet Research*. 2015; 17:e166. [PubMed: 26155878]
73. Monahan KC, Oesterle S, Rhew I, Hawkins JD. The Relation Between Risk and Protective Factors for Problem Behaviors and Depressive Symptoms, Antisocial Behavior, and Alcohol Use in Adolescents. *Journal of Community Psychology*. 2014; 42:621–38. [PubMed: 26985122]
74. Murphy SA, Lynch KG, Oslin D, McKay JR, TenHave T. Developing adaptive treatment strategies in substance abuse research. *Drug & Alcohol Dependence*. 2007; 88:S24–S30. [PubMed: 17056207]
75. Murray, DM. Design and analysis of group-randomized trials. Oxford University Press; 1998.
76. Murray DM, Lee Van Horn M, Hawkins JD, Arthur MW. Analysis strategies for a community trial to reduce adolescent ATOD use: a comparison of random coefficient and ANOVA/ANCOVA models. *Contemporary Clinical Trials*. 2006; 27:188–206. [PubMed: 16324889]
77. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: a review of recent methodological developments. *American Journal of Public Health*. 2004; 94:423–32. [PubMed: 14998806]
78. Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, et al. Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychological Methods*. 2012; 17:457–77. [PubMed: 23025433]

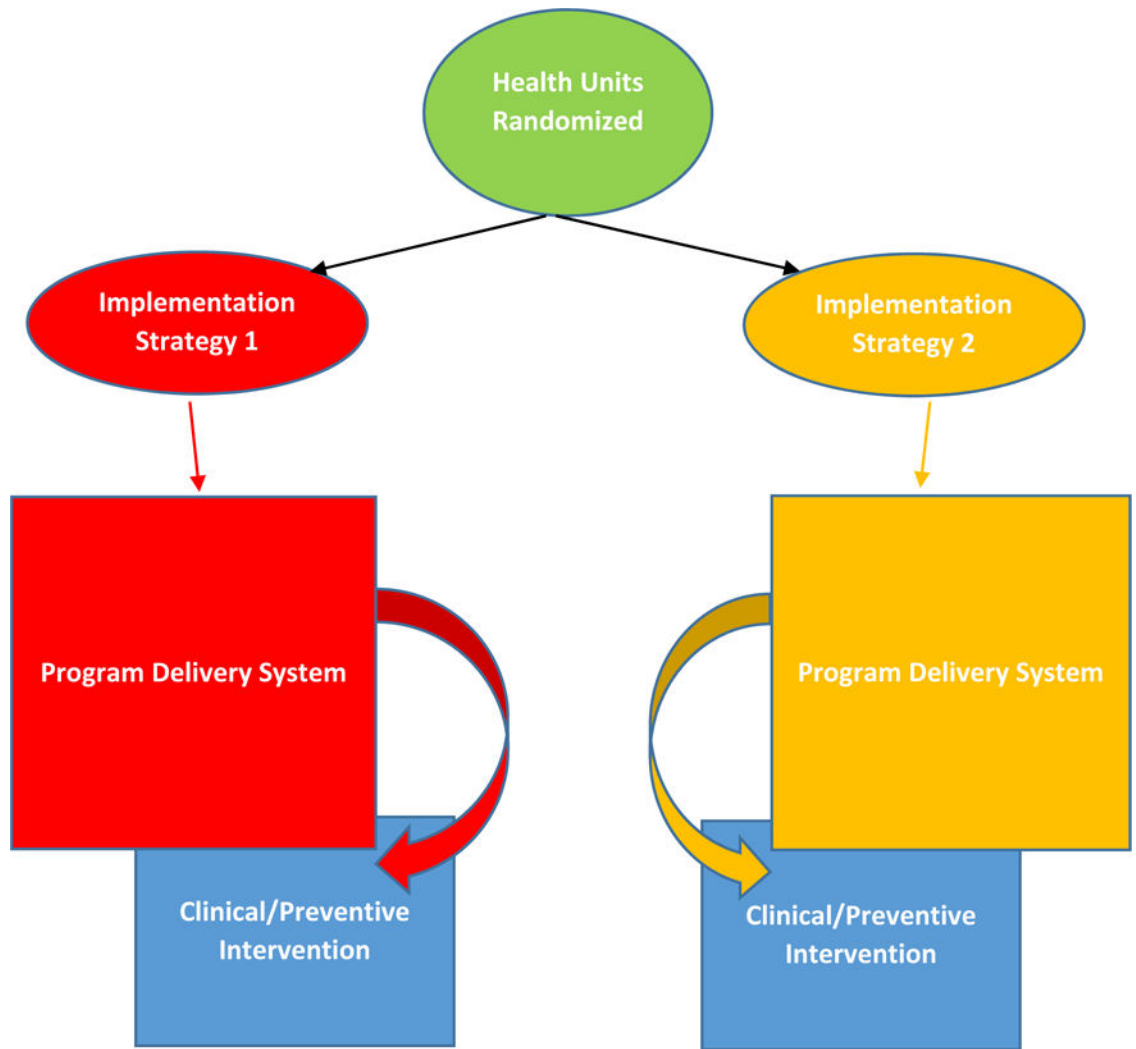
79. National Research Council, Institute of Medicine. Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities. Washington, DC: The National Academies Press; 2009.
80. Neumann MS, Sogolow ED. Replicating effective programs: HIV/AIDS prevention technology transfer. *AIDS Education and Prevention: Official Publication of the International Society for AIDS Education*. 2000; 12:35–48. [PubMed: 11063068]
81. Nilsen P. Making sense of implementation theories, models, and frameworks. *Implementation Science*. 2015; 10
82. Oesterle S, Hawkins JD, Fagan A, Abbott R, Catalano R. Variation in the Sustained Effects of the Communities That Care Prevention System on Adolescent Smoking, Delinquency, and Violence. *Prevention Science*. 2014; 15:138–45. [PubMed: 23412948]
83. Palinkas LA, Aarons GA, Horwitz S, Chamberlain P, Hurlburt M, Landsverk J. Mixed method designs in implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*. 2011; 38:44–53. [PubMed: 20967495]
84. Patterson GR. Interventions for boys with conduct problems: multiple settings, treatments, and criteria. *Journal of Consulting and Clinical Psychology*. 1974; 42:471–81. [PubMed: 4847255]
85. Pellegrini CA, Hoffman SA, Collins LM, Spring B. Optimization of remotely delivered intensive lifestyle treatment for obesity using the Multiphase Optimization Strategy: Opt-IN study protocol. *Contemporary Clinical Trials*. 2014; 38:251–59. [PubMed: 24846621]
86. Perl HI. Addicted to discovery: Does the quest for new knowledge hinder practice improvement? *Addictive Behaviors*. 2011; 11:590–96.
87. Piantadosi, S. *Clinical trials: a methodologic perspective*. New York: Wiley; 1997.
88. Poduska J, Kellam SG, Brown CH, Ford C, Windham A, et al. Study protocol for a group randomized controlled trial of a classroom-based intervention aimed at preventing early risk factors for drug abuse: integrating effectiveness and implementation research. *Implementation Science*. 2009; 4:56. [PubMed: 19725979]
89. Poduska JM, Kellam SG, Wang W, Brown CH, Ialongo NS, Toyinbo P. Impact of the Good Behavior Game, a universal classroom-based behavior intervention, on young adult service use for problems with emotions, behavior, or drugs or alcohol. *Drug and Alcohol Dependence*. 2008; 95:S29–S44. [PubMed: 18249508]
90. Powell BJ, McMillen JC, Proctor EK, Carpenter CR, Griffey RT, et al. A compilation of strategies for implementing clinical innovations in health and mental health. *Medical care research and review: MCRR*. 2012; 69:123–57. [PubMed: 22203646]
91. Prior M, Elouafkaoui P, Elders A, Young L, Duncan E, et al. Evaluating an audit and feedback intervention for reducing antibiotic prescribing behaviour in general dental practice (the RAPiD trial): a partial factorial cluster randomised trial protocol. *Implementation Science*. 2014; 9:50. [PubMed: 24758164]
92. Proctor E, Powell BJ, McMillen JC. Implementation strategies: Recommendations for specifying and reporting. *Implementation Science*. 2013; 8:139–50. [PubMed: 24289295]
93. Purtle J, Peters R, Brownson RC. A review of policy dissemination and implementation research funded by the National Institutes of Health, 2007–2014. *Implementation Science*. 2016; 11
94. Rabin, BA., Brownson, EC. Developing the Terminology for Dissemination and Implementation Research. In: Brownson, RC, Colditz, GA., Proctor, EK., editors. *Dissemination and Implementation Research in Health*. New York: Oxford; 2012. p. 23-51.
95. Raudenbush SW. Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*. 1997; 2:173–85.
96. Reid, JB., Taplin, PS., Lorber, R. A social interactional approach to the treatment of abusive families. In: Stuart, RB., editor. *Violent behavior: Social learning approaches to prediction, management, and treatment*. New York, NY: Brunner/Mazel; 1981. p. 83-101.
97. Rogers, EM. *Diffusion of Innovations*. New York: The Free Press; 1995.
98. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70:41–55.
99. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*. 1974; 66:688–701.

100. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ (Clinical research ed)*. 1996; 312:71–72.
101. Saldana L. The stages of implementation completion for evidence-based practice: protocol for a mixed methods study. *Implementation Science*. 2014; 9:43. [PubMed: 24708893]
102. Shojania KG, Grimshaw JM. Evidence-based quality improvement: the state of the science. *Health Affairs*. 2005; 24:138–50. [PubMed: 15647225]
103. Spoth R, Gyll M, Redmond C, Greenberg M, Feinberg M. Six-year sustainability of evidence-based intervention implementation quality by community-university partnerships: The PROSPER study. *American Journal of Community Psychology*. 2011; 48:412–25. [PubMed: 21394561]
104. Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society*. 2010
105. Szapocznik, J., Duff, JH., Schwartz, SJ., Muir, JA., Brown, CH. Brief Strategic Family Therapy Treatment for Behavior Problem Youth: Theory, Intervention, Research, and Implementation. In: Sexton, T., Lebow, J., editors. *Handbook of Family Therapy: The Science and Practice of Working with Families and Couples*. Abingdon: Routledge; 2015.
106. Thistlethwaite DL, Campbell DT. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of educational psychology*. 1960; 51:309.
107. Twardella D, Brenner H. Effects of practitioner education, practitioner payment and reimbursement of patients' drug costs on smoking cessation in primary care: A cluster randomised trial. *Tobacco Control*. 2007; 16:15–21. [PubMed: 17297068]
108. Valente TW, Palinkas LA, Czaja S, Chu KH, Brown CH. Social network analysis for program implementation. *PLoS ONE*. 2015; 10:e0131712. [PubMed: 26110842]
109. Van Achterberg T, Schoonhoven L, Grol R. Nursing implementation science: how evidence-based nursing requires evidence-based implementation. *Journal of Nursing Scholarship*. 2008; 40:302–10. [PubMed: 19094144]
110. Weiss CH, Poncela-Casasnovas J, Glaser JJ, Pah AR, Persell SD, et al. Adoption of a High-Impact Innovation in a Homogeneous Population. *Physical Review X*. 2014; 4:041008. [PubMed: 25392742]
111. Wyman PA, Henry D, Knoblauch S, Brown CH. Designs for testing group-based interventions with limited numbers of social units: The dynamic wait-listed and regression point displacement designs. *Prevention Science*. 2015; 16:956–66. [PubMed: 25481512]
112. Wyrick DL, Rulison KL, Fearnow-Kenney M, Milroy JJ, Collins LM. Moving beyond the treatment package approach to developing behavioral interventions: addressing questions that arose during an application of the Multiphase Optimization Strategy (MOST). *Translational Behavioral Medicine*. 2014; 4:252–59. [PubMed: 25264465]

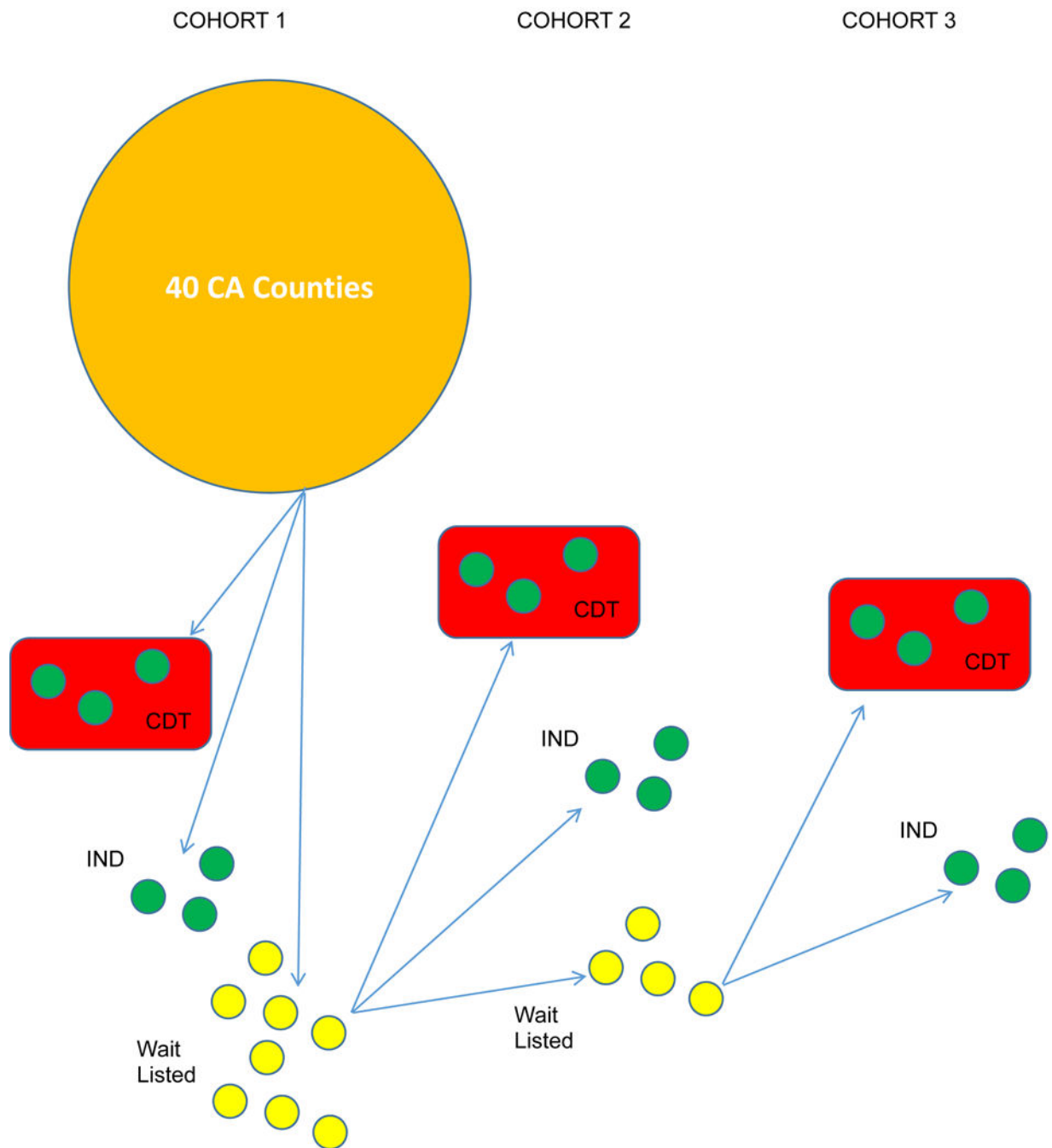


**Figure 1. Traditional Translational Pipeline**

\*These dissemination and implementation stages include systematic monitoring, evaluation, and adaptation as required.



**Figure 2.** Focus of Research in a Head-to-Head Randomized implementation Trial with Identical Clinical/Preventive Intervention and Different Implementation Strategies



**Figure 3.** Randomize 40 Counties in CA to Independent County (IND) or Community Development Team (CDT) Implementation Strategy and Time (Cohort) Using a Randomized Roll-Out Design; 11 Counties in OH were Separately Randomized in a Fourth Cohort to the Same Implementation Strategies.



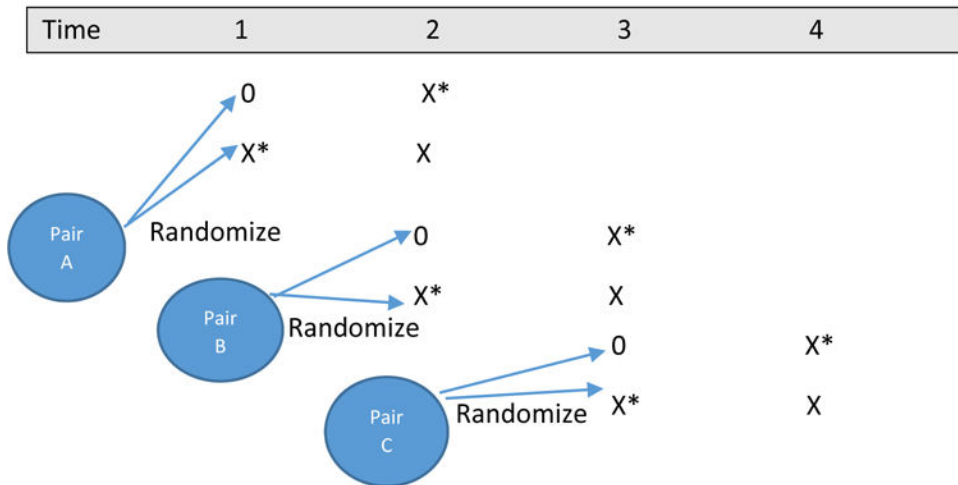
(a)

Time	1	2	3	4
Cohort A	0	X*	X	X
Cohort B	0	0	X*	X
Cohort C	0	0	0	X*

(b)

Time	1	2	3	4
Cohort A	0	X*	X	X
	0	Y*	Y	Y
Cohort B	0	0	X*	X
	0	0	Y*	Y
Cohort C	0	0	0	X*
	0	0	0	Y*

(c)



**Figure 4. Roll-Out Randomized Designs**

Symbol “O” refers to implementation as usual, “X\*” and “Y\*” are introductions on new implementation strategies, and “X” and “Y” are continuation of these strategies over extended periods of time.