An Overview of Research and Evaluation Designs for Dissemination and Implementation

C. Hendricks Brown,[1] Geoffrey Curran,[2]
Lawrence A. Palinkas,[3] Gregory A. Aarons,[4]
Kenneth B. Wells,[5] Loretta Jones,[6] Linda M. Collins,[7]
Naihua Duan,[8] Brian S. Mittman,[9] Andrea Wallace,[10]
Rachel G. Tabak,[11] Lori Ducharme,[12]
David A. Chambers,[13] Gila Neta,[13] Tisha Wiley,[14]
John Landsverk,[15] Ken Cheung,[16]
and Gracelyn Cruden[1,17]

[1]Department of Psychiatry and Behavioral Sciences, Feinberg School of Medicine,
Northwestern University, Chicago, Illinois 60611; email: hendricks.brown@northwestern.edu

[2]Division of Health Services Research, Psychiatric Research Institute, University of Arkansas
for Medical Sciences, Little Rock, Arkansas 72205; email: currangeoffreym@uams.edu

[3]Department of Children, Youth and Families, School of Social Work, University of Southern
California, Los Angeles, California 90089; email: palinkas@usc.edu

[4]Department of Psychiatry, University of California, San Diego, School of Medicine, La Jolla,
California 92093; email: gaarons@ucsd.edu

[5]Center for Health Services and Society, Semel Institute for Neuroscience and Human
Behavior, University of California, Los Angeles, California 90024;
email: KWells@mednet.ucla.edu

[6]Healthy African American Families, Los Angeles, California 90008;
email: lorettajones@haafii.org

[7]The Methodology Center and Department of Human Development & Family Studies,
Pennsylvania State University, University Park, Pennsylvania 16802; email: lmcollins@psu.edu

[8]Department of Psychiatry, Columbia University Medical Center, Columbia University, New
York, NY 10027; email: naihuaduan@gmail.com

[9]VA Center for Implementation Practice and Research Support, Virginia Greater Los Angeles
Healthcare System, North Hills, California 91343; email: Brian.Mittman@va.gov

[10]College of Nursing, The University of Iowa, Iowa City, Iowa 52242;
email: andrea-wallace@uiowa.edu

[11]Prevention Research Center, George Warren Brown School, Washington University,
St. Louis, Missouri 63105; email: rtabak@wustl.edu

[12]National Institute of Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda,
Maryland 20814; email: lori.ducharme@nih.gov

[13]Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville,
Maryland 20850; email: dchamber@mail.nih.gov, netagil@mail.nih.gov

[14]National Institute on Drug Abuse, National Institutes of Health, Bethesda, Maryland 20814;
email: wileytr@nida.nih.gov

[15]Oregon Social Learning Center, Eugene, Oregon 97401; email: jlandsverk@aol.com

[16]Mailman School of Public Health, Columbia University, New York, NY 10032;
email: kencheung2@gmail.com

[17]Department of Health Policy and Management, University of North Carolina, Chapel Hill,
North Carolina 27514; email: gcruden@live.unc.edu

Abstract

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. This article is one product of a design workgroup that was 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. This article emphasizes randomized and nonrandomized 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.

INTRODUCTION

Medicine and public health have made great progress using rigorous randomized clinical trials to determine whether an intervention is efficacious. The standards set by Fisher (40), 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 the comparative effectiveness of one active intervention against another. Investigators have made practical modifications to the individual-level randomized clinical trial to test for program or intervention effectiveness under realistic conditions in randomized field trials (41), as well as for interventions delivered at the group level (77), for multilevel interventions (14, 15), for complex (35, 70) or multiple component interventions (32), and for tailoring or adapting the intervention to a subject's response to targeted outcomes (32, 74) or to different social, physical, or virtual environments (13). To test efficacy or effectiveness, researchers now have a large family of designs 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). Although 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-methods designs, reasonable consensus has grown across diverse fields about when certain designs should be used and which sample size requirements and design protocols are necessary to maximize internal validity (43, 87).

Dissemination and implementation research represents a distinct stage, and the designs for this newer field of research are currently not as well established as are those for efficacy

and effectiveness. A lack of understanding of 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 types of interventions can be delivered in and function effectively in varying 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). We refer to these as the 7 Ps.

This article uses the term clinical/preventive intervention to refer to a single set or multiple sets of these 7 Ps, which 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, whereas implementation involves "the use of strategies to adopt and integrate evidence-based health interventions and change practice patterns within specific settings" (49, p. 1275). 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 afterward.

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 (21, 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 about whether and to what extent designs involving randomized assignment should be used in dissemination and implementation studies (79), as well as about the relative contributions of qualitative, quantitative, and mixed methods in dissemination and implementation designs (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, 26, 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 (28). Debates about design also revolve around conflicting views pertaining to the underlying scientific issue of how much emphasis to place on internal validity compared with external validity (52).
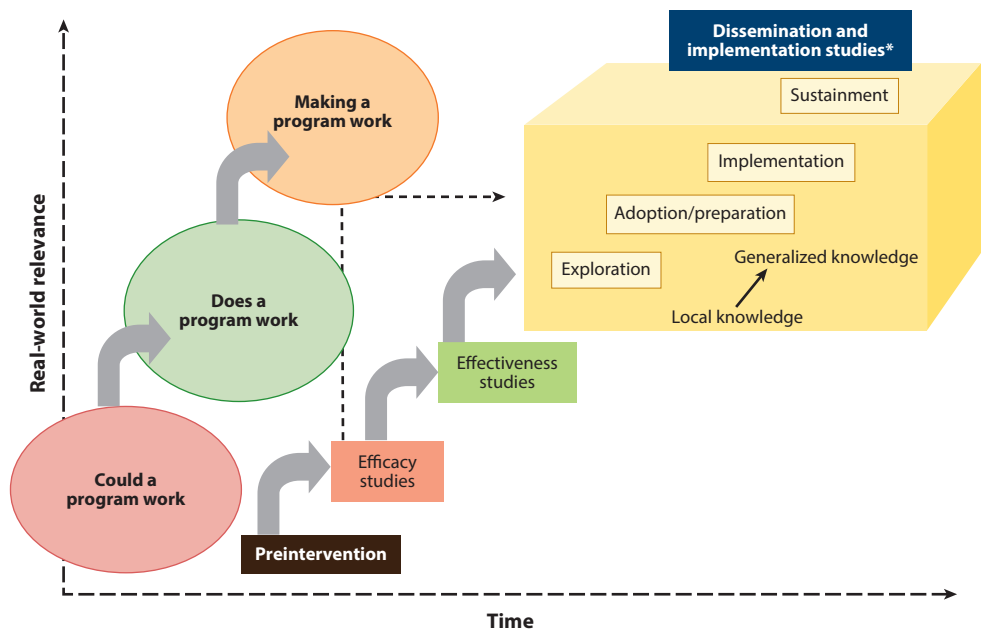
In this article, we introduce a conceptual view of the traditional translational pipeline that was formulated as a continuum of research originally known as Levy's arrow (68). This traditional translational pipeline is commonly used by the National Institutes of Health (NIH) and other research-focused organizations to move scientific knowledge from basic and other preintervention research to efficacy and effectiveness trials and to a stage that reaches the public (66, 79). By no means does all dissemination and implementation research follow this traditional translational pipeline, so we mention in the discussion section three different classes of research design 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 considered.

This article is a product of a design workgroup formed in 2013 by the NIH to address dissemination and implementation research. We established a shared definition of terms, which required

significant compromise because 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 number of individuals in 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. Throughout this article, we provide other consensus definitions involving dissemination and implementation as well as statistical design terms.

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 preintervention 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 their impact under ideal conditions. A highly trained research team would typically deliver this program to a homogeneous group of subjects with careful monitoring and supervision to ensure high fidelity in this efficacy stage. Efficacy trials can answer only questions of whether a clinical/preventive intervention could work under rigorous conditions; therefore, such a program or practice that

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

Figure 1

Traditional translational pipeline from preintervention, efficacy, effectiveness, and dissemination and implementation studies.

demonstrates sufficient efficacy would then be followed, in the traditional research pipeline, by the next stage, 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. In these effectiveness trials, clinicians, other practitioners, or trained individuals from the community typically deliver the clinical/preventive intervention with ongoing supervision by researchers. Also, in contrast with the homogeneous group of subjects used in efficacy trials, a more heterogeneous group of study participants is generally included in effectiveness trials. 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. According to this pipeline, the clinical/preventive intervention must have already demonstrated effectiveness before an implementation study can be conducted. 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 reexamine effectiveness in the midst of an implementation research design (39, 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 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 a population-level impact, it must not only be an effective program, but also reach a large portion of the population, be delivered with fidelity, and be maintained (50). Within the dissemination and implementation research agenda, researchers have distinguished some phases of the implementation process itself. A common exemplar, the EPIS conceptual model of the implementation process (1), identifies four phases: exploration, preparation, implementation, and sustainment, 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, given its 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 stakeholders decide to adopt it. 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 (36). The final phase is sustainment and refers to how host delivery systems and organizations maintain or extend the supports as well as the clinical/preventive intervention, especially after the initial funding period 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 article as the implementation strategy, which is viewed as distinct from, but generally dependent on, the specific clinical/preventive intervention that is being adopted.

Figure 1 also contrasts local formative evaluation or quality improvement with 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 translational 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; in fact, the sequential progression of intervention studies is often cyclical (14). Trial designs may change through this pipeline, as well. Efficacy or effectiveness trials nearly always use random assignment, whereas implementation research often requires trade-offs 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 that provide within-site, between-site, and within- and between-site comparisons of implementation strategies.

Within-Site Designs

Within-site designs can be used to evaluate implementation successes or failures by examining changes that occur inside an organization, community, or system. They can be comparatively simple and inexpensive, as in the example we provide for post designs, or vary in complexity and expense in pre-post and multiple baseline 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 one example, consider the introduction of rapid oral HIV testing in a hospital-based dental clinic, a site that may be useful from a population health standpoint, based on access to a population that may include individuals who have not been tested and on the convenience, speed, sensitivity, and specificity of this test. The Centers for Disease Control and Prevention (CDC) recently proposed that dental clinics could deliver this new technology, and public health questions remain 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 phase). Here, a key process measure is the rate at 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 detection of subjects who did not know they were HIV positive. 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 to predict program adoption. For example, all 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 (24). 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 it was being used in any California communities (26). 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 preimplementation levels. Some clinical/preventive interventions are already being used in organizations and communities, but they 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 or revised implementation strategy for a clinical/preventive intervention that is already being used; process or output is measured prior to and after the new implementation strategy begins. Effects due to the new implementation strategy are inferred by comparing pre to post changes within this site. 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 given 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 the publication of guidelines (8). The effects of a black box warning on antidepressant prescriptions for depressed youth were evaluated by comparing prescription rates (46) and adverse event reports (47) prior to and after introduction of the warning.

A variant of the pre-post design involves a multiple-baseline time-series design. After an outlet store reward and reminder system was implemented, Biglan and colleagues (7) examined the prevalence of tobacco being sold to young people over multiple time points without personnel checking birthdays. Tracking this prevalence over time can examine whether the reward and reminder system has sustained effects.

Between-Site Designs

In contrast with previous designs that examined changes over time within one or more sites that were exposed to the same dissemination or implementation strategy, the designs in this section compare processes and output among sites that have 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 (54) 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. 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 counties that receive this policy intervention compared with those that do not.

One illustration where self-selection is important comes from an educational outreach strategy to affect physicians' prescribing practices for 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, which severely limited 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). Interpreting 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 us understand why some organizations adopt an implementation strategy and 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 with what now exists (42). In this trial, the sites (practices, communities, organizations) are assigned randomly to the new active implementation or usual-practice condition. Because random assignment occurs at the group rather than individual level, this method forms a cluster-randomized design (75, 95). Process and output measures used as the primary end points 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 either to receive a supportive strategy for implementing a combination of evidence-based school and family interventions to prevent youth substance abuse or to maintain 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 randomizing larger health organizational sites, investigators can sometimes randomize smaller units within each organization. Randomization could occur at the level of the ward, team, or 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 the site as a blocking factor in contrast with 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 to 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, then 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, feedback was given at ward-level meetings (44). 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. Individual-level randomized implementation designs can be appropriate provided that (*a*) leakage of the implementation to other patients is minimal and (*b*) the implementation's impact is not attenuated as a result of only a portion of patients being exposed to it. One 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. In 2013, Aspry et al. (4) reviewed system, practitioner, and individual patient-level randomized trials to evaluate level-specific lipid management approaches that all included a health information technology component, and they concluded that system and individual patient implementations showed better lipid management than did practitioner-level information technologies.

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 does. The effect of providing incentives or supports for succeeding to implement or penalties for failing to implement can be evaluated with such a design. One example involves the pay-for-performance (P4P) strategy to increase therapist adherence to a protocol for adolescent substance abusers (45); in this scenario, 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. With this design, the same clinical/preventive intervention is used for both arms of the 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).

One example of the use of such a head-to-head randomized implementation trial is the CAL-OH trial, which compares two alternative strategies to implement multidimensional treatment foster care (MTFC) at the county level (24, 27). MTFC is an evidence-based program for foster children and their families and 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). **Figure 3** shows a diagram of the trial design for 40 California counties. 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 rollout 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 (25) was developed and used to evaluate the implementation process, including the quality of preparedness and training to deliver MTFC, the speed at which 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).
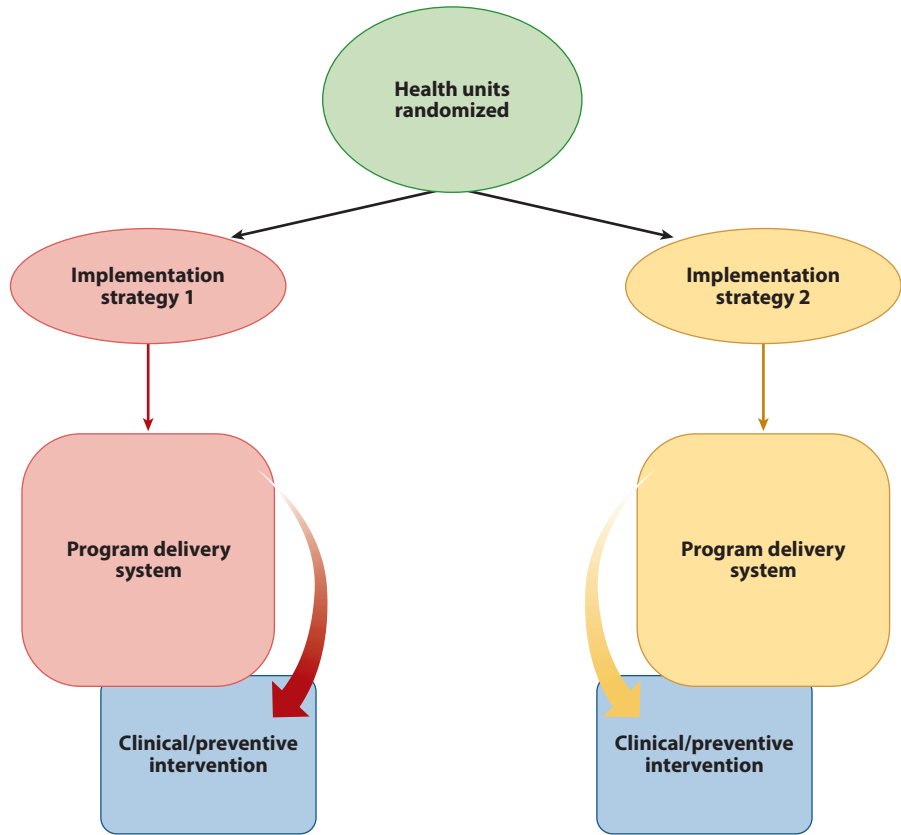
Figure 2

Focus of research in a head-to-head randomized implementation trial with identical clinical/preventive intervention and different implementation strategies.

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. One 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. One 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 has focused on a single evidence-based clinical/preventive intervention. This one-choice

Cohort 1 **Cohort 2** **Cohort 3**

- ● Community dependent team
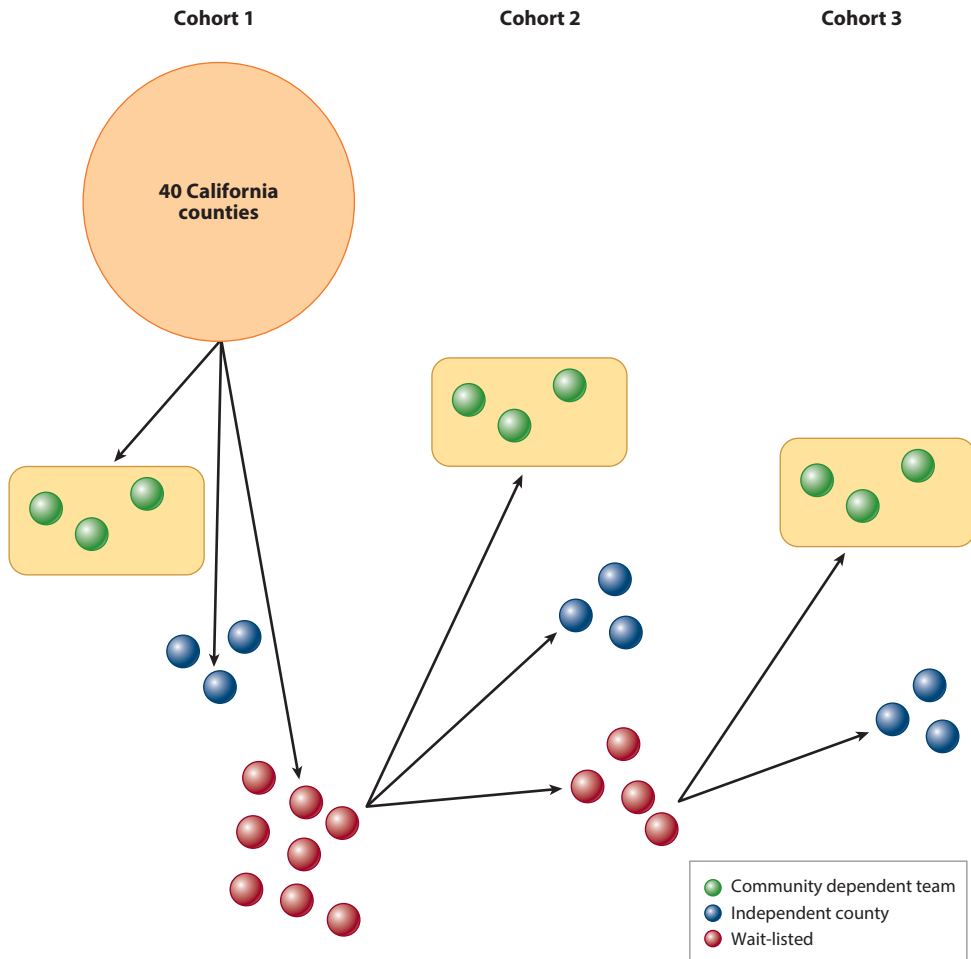- ● Independent county
- ● Wait-listed

Figure 3

Design to assign 40 counties in California to an independent county or community development team implementation strategy and time (cohort) using a randomized rollout design; 11 counties in Ohio were separately randomized in a fourth cohort to the same two implementation strategies (not shown).

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.

One example of a randomized implementation trial of such a decision support system for the 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) (56) comprehensive community support system against the community control condition, under which community leaders only received information about their community's risk and protective factor profile but were given 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 as well as drug use and violence outcomes at the community level (19, 57, 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, and 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 of their interaction. One example of a 3×3 design is the evaluation of three alternative alcohol abuse screening tools for use in emergency departments, which are deployed 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 (34).

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 by communities or organizations as unethical to withhold both strategies; thus, units may be assigned to either implementation or to 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) (30, 33, 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, and the clinician level, as well as key processes: planning, educating, financing, restructuring, management of quality and/or policy (90). Many of these components are thought to be necessary for an implementation strategy to work properly. Because these implementation components can be specified (92) and connected (105) in diverse ways, and because these approaches vary 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 (31, 33, 38). The set of components that best meets the optimization criterion is identified on the basis of the trial's results. In the third phase, evaluation, a standard randomized implementation trial is conducted that compares the optimized intervention against an appropriate comparison condition.

One example of a MOST implementation study involved determining which of a set of three components improved the fidelity of teacher delivery of Healthwise (22, 23) 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; 56 schools were assigned to one of 8 experimental conditions (23).

The sequential multiple assignment randomized implementation trial (SMART) implementation design is a special case of the factorial experiment (33, 67, 78) that involves multistage randomizations where the site-level implementation process can be modified if unsuccessful. Such an adaptive approach to enhancing implementation can optimize allocation of available resources

(32) 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 their use of evidence-based interventions (65). In a subsequent study in 2014, Kilbourne et al. used SMART to examine an adaptive version of REP (64). Initially, 80 community-based outpatient clinics are provided with the original REP implementation strategy. 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 rerandomized 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. (29) propose a SMART design with adaptive randomization (SMART-AR) in which the randomization probabilities of the strategies are updated on the basis of outcomes in other sites, so as to improve the expected outcome. Such adaptations are potentially infinite if the interventions are complex and the intended populations and settings are 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, two experimental implementation factors are directed toward 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. One example of this design is a trial for smoking cessation implementation to test 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 rollout randomized implementation trial to refer to the broad class of designs with which the timing of the start of an implementation strategy is randomly assigned. In the simplest rollout 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, investigators conduct a between-units comparison of those units that are assigned to the new implementation strategy and those that remain in 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 4a shows this simplest type of rollout design. 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., the point at which 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 rollout 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 are often willing to accept this type of rollout design over a traditional design when there are obvious or perceived advantages to receiving the new implementation strategy or when it is unethical to withhold a new implementation strategy throughout the study (17).
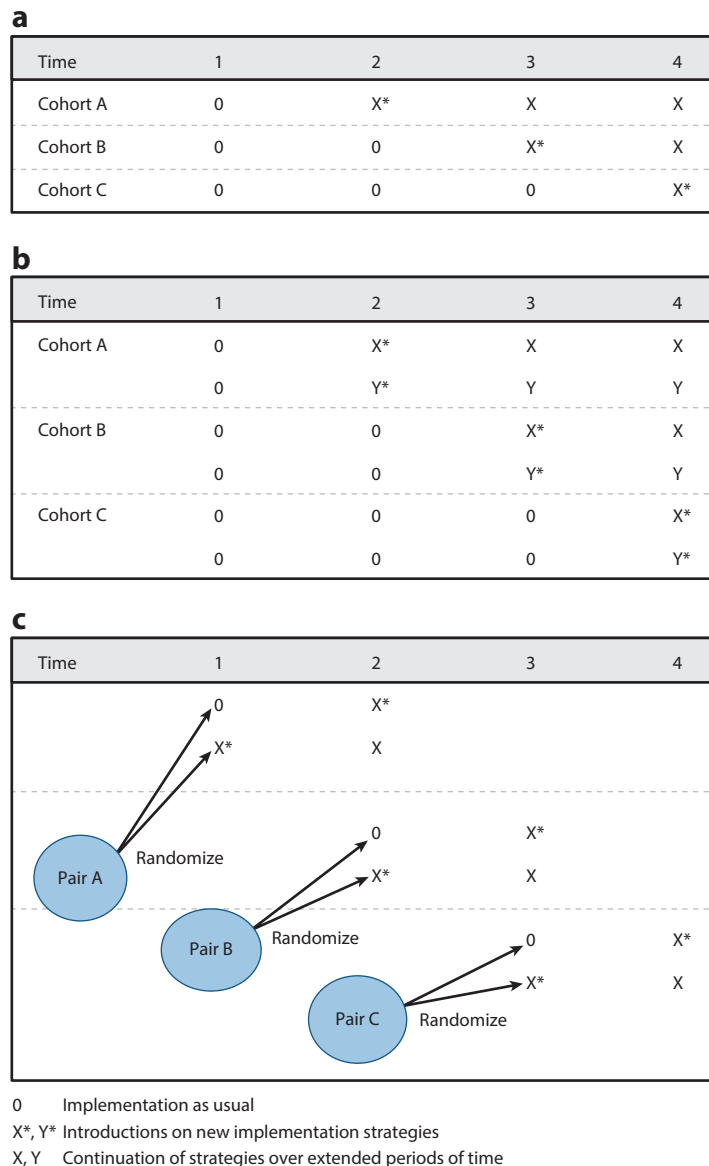
a

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

Time	1	2	3	4

0   Implementation as usual
X*, Y* Introductions on new implementation strategies
X, Y   Continuation of strategies over extended periods of time

Figure 4

Schematics of three rollout randomized designs that determine the timing of changes from usual practice, startup or continuation of one or more implementation strategies.

Other rollout designs can be used for implementation (111). In **Figure 4b**, 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*. 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, which is conducted by the National Institute on Drug Abuse (NIDA) and its colleagues (6). We refer to this strategy as a head-to-head rollout trial to distinguish it from the stepped-wedge designs discussed above.

In addition, other types of rollout designs could be used; to our knowledge, however, they have not yet been used in implementation. A pairwise enrollment rollout design (111) differs from the two we previously discussed and is similar to the original design for a rollout randomized effectiveness trial of an HIV community-based intervention. In the 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 subsequent years so that over time a true randomized trial with sufficient numbers of units could be conducted. This type of pairwise enrollment rollout design, shown in **Figure 4c**, could also work for implementation trials, thereby eliminating the need to enumerate all sites at the beginning of the study.

DISCUSSION

In this article, 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 interventions 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 have presented designs for this pipeline, including within-site comparisons alone, between-site comparisons, and within- and between-site comparisons with rollout designs. A wide range of factorial designs can also be used to evaluate multiple implementation components. Many of the designs we have described involve randomization, which can often strengthen inferences; however, in many situations, randomization is not possible or required nor even advisable. For example, in some instances, policies are disseminated or implemented by law (93) or through another 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).

Many excellent dissemination and implementation designs address issues beyond those relevant to the traditional translational pipeline, and our focus on pipeline designs should in no way be interpreted as minimizing the importance of these other designs. We note three broad types not discussed here: designs where effectiveness of the clinical preventive intervention is to be evaluated alongside questions that involve implementation, which include hybrid designs (37) and continuously evolving interventions (72); designs that address quality improvement for local knowledge (28); and designs that involve simulation or synthetic experiments in dissemination and implementation. These designs are presented elsewhere.

We did not provide details on statistical power in this review, despite its obvious importance. We do make a few general comments on power. Dissemination and implementation research, because it often tests system-level strategies, generally requires multilevel data with sizeable numbers of groups (e.g., practices or organizations) as well as individuals, because the statistical power of such designs is influenced most strongly by the number of units at the highest (group) 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-methods designs is another approach to addressing limited statistical power (83).

We also have not 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). Second, 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, 26, 101).

In terms of causality, nonrandomized designs generally provide less confidence that the numerical comparisons are due to the differences in implementation conditions as compared with randomized designs. 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 design maintains some protection. Such designs cannot completely rule out all unmeasured external effects in the way that randomized trials can. Some subtle issues can also 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 that involve 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 because 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 can address the 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.

Although we have provided some illustrations of the use of many of the designs discussed above, this article does not provide recommendations regarding the appropriateness of one design over another. The choice of design is a complex process that requires 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 these factors are beyond the scope of this review. 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 significant changes in organizations and communities; as such, community leaders, organizational leaders, and policy makers have far more at stake than do the evaluators. The most attractive scientific design on paper will not happen without the endorsement and agreement of the communities and organizations where these strategies are implemented.

DISCLOSURE STATEMENT

ACKNOWLEDGMENTS

LITERATURE CITED

1. Aarons GA, Hurlburt M, Horwitz SM. 2011. Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Adm. Policy Ment. Health Ment. Health Serv. Res.* 38:4–23

2. Aarons GA, Sommerfeld DH, Walrath-Greene CM. 2009. Evidence-based practice implementation: the impact of public versus private sector organization type on organizational support, provider attitudes, and adoption of evidence-based practice. *Implement. Sci.* 4:83

3. Angrist JD, Imbens GW, Rubin DB. 1996. Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* 91:444–55

4. Aspry KE, Furman R, Karalis DG, Jacobson TA, Zhang AM, et al. 2013. Effect of health information technology interventions on lipid management in clinical practice: a systematic review of randomized controlled trials. *J. Clin. Lipidol.* 7:546–60

5. Baer DM, Wolf MM, Risley TR. 1968. Some current dimensions of applied behavior analysis. *J. Appl. Behav. Anal.* 1:91–97

6. Belenko S, Wiley T, Knight D, Dennis M, Wasserman G, Taxman F. 2015. A new behavioral health services cascade framework for measuring unmet addiction health services needs and adolescent offenders: conceptual and measurement challenges. *Addict. Sci. Clin. Pract.* 10(Suppl. 1):A4

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

8. Blackstock OJ, King JR, Mason RD, Lee CC, Mannheimer SB. 2010. Evaluation of a rapid HIV testing initiative in an urban, hospital-based dental clinic. *AIDS Patient Care STDs* 24:781–85

9. Bloom HS. 1984. Accounting for no-shows in experimental evaluation designs. *Eval. Rev.* 8:225–46

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

11. Brown CH, Chamberlain P, Saldana L, Padgett C, Wang W, Cruden G. 2014. 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. *Implement. Sci.* 9:134

12. Brown CH, Liao J. 1999. Principles for designing randomized preventive trials in mental health: an emerging developmental epidemiology paradigm. *Am. J. Commun. Psychol.* 27:673–710

13. Brown CH, Mohr DC, Gallo CG, Mader C, Palinkas L, et al. 2013. A computational future for preventing HIV in minority communities: how advanced technology can improve implementation of effective programs. *J. Acquir. Immune Defic. Syndr.* 63:S72–84

14. Brown CH, Ten Have TR, Jo B, Dagne G, Wyman PA, et al. 2009. Adaptive designs for randomized trials in public health. *Annu. Rev. Public Health* 30:1–25

15. Brown CH, Wang W, Kellam SG, Muthén BO, Petras H, et al. 2008. 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 Alcohol Depend.* 95:S74–104

16. Brown CH, Wyman PA, Brinales JM, Gibbons RD. 2007. The role of randomized trials in testing interventions for the prevention of youth suicide. *Int. Rev. Psychiatry* 19:617–31

17. 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

18. Brown EC, Hawkins JD, Arthur MW, Briney JS, Abbott RD. 2007. Effects of communities that care on prevention services systems: findings from the Community Youth Development Study at 1.5 years. *Prev. Sci.* 8:180–91

19. Brown EC, Hawkins JD, Rhew IC, Shapiro VB, Abbott RD, et al. 2014. Prevention system mediation of communities that care effects on youth outcomes. *Prev. Sci.* 15:623–32

20. Brownson RC, Colditz GA, Proctor EK, eds. 2012. *Dissemination and Implementation Research in Health: Translating Science to Practice*. London: Oxford Univ. Press

21. Brownson RC, Diez Roux AV, Swartz K. 2014. Commentary: Generating rigorous evidence for public health: the need for new thinking to improve research and practice. *Annu. Rev. Public Health* 35:1–7

22. Caldwell LL, Patrick ME, Smith EA, Palen L-A, Wegner L. 2010. Influencing adolescent leisure motivation: intervention effects of HealthWise South Africa. *J. Leis. Res.* 42:203–20

23. Caldwell LL, Smith EA, Collins LM, Graham JW, Lai M, et al. 2012. Translational research in South Africa: evaluating implementation quality using a factorial design. *Child Youth Care Forum* 41(2):119–36

24. Chamberlain P. 2003. *Treating Chronic Juvenile Offenders: Advances Made Through the Oregon Multidimensional Treatment Foster Care Model*. Washington, DC: Am. Psychol. Assoc.

25. Chamberlain P, Brown CH, Saldana L. 2011. Observational measure of implementation progress in community based settings: the Stages of Implementation Completion (SIC). *Implement. Sci.* 6:116

26. Chamberlain P, Brown CH, Saldana L, Reid J, Wang W, et al. 2008. Engaging and recruiting counties in an experiment on implementing evidence-based practice in California. *Adm. Policy Ment. Health* 35:250–60

27. Chamberlain P, Price J, Leve LD, Laurent H, Landsverk JA, Reid JB. 2008. Prevention of behavior problems for children in foster care: outcomes and mediation effects. *Prev. Sci.* 9:17–27

28. Cheung K, Duan N. 2013. Design of implementation studies for quality improvement programs: an effectiveness-cost-effectiveness framework. *Am. J. Public Health* 104:e23–30

29. Cheung YK, Chakraborty B, Davidson KW. 2015. Sequential Multiple Assignment Randomized Trial (SMART) with adaptive randomization for quality improvement in depression treatment program. *Biometrics* 71:450–59

30. Collins LM, Baker TB, Mermelstein RJ, Piper ME, Jorenby DE, et al. 2011. The multiphase optimization strategy for engineering effective tobacco use interventions. *Ann. Behav. Med.* 41:208–26

31. Collins LM, Dziak JJ, Li R. 2009. Design of experiments with multiple independent variables: a resource management perspective on complete and reduced factorial designs. *Psychol. Methods* 14:202–24

32. Collins LM, Murphy SA, Bierman KL. 2004. A conceptual framework for adaptive preventive interventions. *Prev. Sci.* 5:185–96

33. Collins LM, Nahum-Shani I, Almirall D. 2014. Optimization of behavioral dynamic treatment regimens based on the Sequential, Multiple Assignment, Randomized Trial (SMART). *Clin. Trials* 11:426–34

34. Coulton S, Perryman K, Bland M, Cassidy P, Crawford M, et al. 2009. Screening and brief interventions for hazardous alcohol use in accident and emergency departments: a randomised controlled trial protocol. *BMC Health Serv. Res.* 9:114

35. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, et al. 2008. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ (Clin. Res. Ed.)* 337:a1655

36. Cross W, West J, Wyman PA, Schmeelk-Cone K, Xia Y, et al. 2015. Observational measures of implementer fidelity for a school-based prevention intervention: development, reliability, and validity. *Prev. Sci.* 16:122–32

37. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. 2012. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Med. Care* 50:217–26

38. Dziak JJ, Nahum-Shani I, Collins LM. 2012. Multilevel factorial experiments for developing behavioral interventions: power, sample size, and resource considerations. *Psychol. Methods* 17:153–75

39. Elliott DS, Mihalic S. 2004. Issues in disseminating and replicating effective prevention programs. *Prev. Sci.* 5:47–53

40. Fisher RAS. 1935. *The Design of Experiments*. Edinburgh: Oliver and Boyd

41. Flay BR. 1986. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Prev. Med.* 15:451–74

42. Folks B, LeBlanc WG, Staton EW, Pace WD. 2011. Reconsidering low-dose aspirin therapy for cardiovascular disease: a study protocol for physician and patient behavioral change. *Implement. Sci.* 6:65

43. Friedman LM, Furberg C, DeMets DL. 1996. *Fundamentals of Clinical Trials*. St. Louis: Mosby-Year Book

44. Fuller C, Michie S, Savage J, McAteer J, Besser S, et al. 2012. The Feedback Intervention Trial (FIT)—improving hand-hygiene compliance in UK healthcare workers: a stepped wedge cluster randomised controlled trial. *PLOS ONE* 7:e41617

45. Garner BR, Godley SH, Dennis ML, Hunter BD, Bair CML, Godley MD. 2012. Using pay for performance to improve treatment implementation for adolescent substance use disorders: results from a cluster randomized trial. *Arch. Pediatr. Adolesc. Med.* 166:938–44

46. Gibbons RD, Brown CH, Hur K, Marcus SM, Bhaumik DK, et al. 2007. Early evidence on the effects of regulators' suicidality warnings on SSRI prescriptions and suicide in children and adolescents. *Am. J. Psychiatry* 164:1356–63

47. Gibbons RD, Segawa E, Karabatsos G, Amatya AK, Bhaumik DK, et al. 2008. Mixed-effects Poisson regression analysis of adverse event reports: the relationship between antidepressants and suicide. *Stat. Med.* 27:1814–33

48. Glasgow RE, Magid DJ, Beck A, Ritzwoller D, Estabrooks PA. 2005. Practical clinical trials for translating research to practice: design and measurement recommendations. *Med. Care* 43:551–57

49. Glasgow RE, Vinson C, Chambers D, Khoury MJ, Kaplan RM, Hunter C. 2012. National Institutes of Health approaches to dissemination and implementation science: current and future directions. *Am. J. Public Health* 102:1274–81

50. Glasgow RE, Vogt TM, Boles SM. 1999. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am. J. Public Health* 89:1322–27

51. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, et al. 2010. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *N. Engl. J. Med.* 363:2587–99

52. Green LW, Glasgow RE. 2006. Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology. *Eval. Health Prof.* 29:126–53

53. Greenhalgh T, Robert G, MacFarlane F, Bate P, Kyriakidou O. 2004. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q.* 82:581–629

54. Hahn EJ, Rayens MK, Butler KM, Zhang M, Durbin E, Steinke D. 2008. Smoke-free laws and adult smoking prevalence. *Prev. Med.* 47:206–9

55. Hall L, Eccles M, Barton R, Steen N, Campbell M. 2001. Is untargeted outreach visiting in primary care effective? A pragmatic randomized controlled trial. *J. Public Health Med.* 23:109–13

56. Hawkins JD, Catalano RF. 2002. *Investing in Your Community's Youth: An Introduction to the Communities that Care System*. South Deerfield, MA: Channing Bete

57. Hawkins JD, Oesterle S, Brown EC, Abbott RD, Catalano RF. 2014. Youth problem behaviors 8 years after implementing the Communities that Care prevention system: a community-randomized trial. *JAMA Pediatrics* 168:122–29

58. Hill ABS. 1961. *Principles of Medical Statistics*. London: Lancet

59. Jabbour M, Curran J, Scott SD, Guttman A, Rotter T, et al. 2013. Best strategies to implement clinical pathways in an emergency department setting: study protocol for a cluster randomized controlled trial. *Implement. Sci.* 8:55

60. Jo B, Asparouhov T, Muthen BO, Ialongo NS, Brown CH. 2008. Cluster randomized trials with treatment noncompliance. *Psychol. Methods* 13:1–18

61. Katz D, Vander Weg M, Fu S, Prochazka A, Grant K, et al. 2009. A before-after implementation trial of smoking cessation guidelines in hospitalized veterans. *Implement. Sci.* 4:58

62. Kegeles SM, Hays RB, Pollack LM, Coates TJ. 1999. Mobilizing young gay and bisexual men for HIV prevention: a two-community study. *AIDS* 13:1753–62

63. Kessler R, Glasgow RE. 2011. A proposal to speed translation of healthcare research into practice: Dramatic change is needed. *Am. J. Prev. Med.* 40:637–44

64. Kilbourne AM, Almirall D, Eisenberg D, Waxmonsky J, Goodrich DE, et al. 2014. 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. *Implement. Sci.* 9:132

65. Kilbourne AM, Neumann MS, Pincus HA, Bauer MS, Stall R. 2007. Implementing evidence-based interventions in health care: application of the replicating effective programs framework. *Implement. Sci.* 2:42

66. Landsverk J, Brown CH, Chamberlain P, Palinkas L, Ogihara M, et al. 2012. Design and analysis in dissemination and implementation research. See Ref. 20, pp. 225–60

67. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. 2012. A "SMART" design for building individualized treatment sequences. *Annu. Rev. Clin. Psychol.* 8:21–48

68. Levy RI. 1982. The National Heart, Lung, and Blood Institute: overview 1980: the Director's report to the NHLBI advisory council. *Circulation* 65:217–25

69. McClure JB, Derry H, Riggs KR, Westbrook EW, St John J, et al. 2012. Questions About Quitting (Q^2): design and methods of a multiphase optimization strategy (MOST) randomized screening experiment for an online, motivational smoking cessation intervention. *Contemp. Clin. Trials* 33:1094–102

70. Med. Res. Counc. (MRC) Health Serv. Public Health Res. Board. 2000. *A Framework for the Development and Evaluation of RCTs for Complex Interventions to Improve Health*. London: MRC. **https://www.mrc.ac.uk/documents/pdf/rcts-for-complex-interventions-to-improve-health/**

71. Mercer SL, DeVinney BJ, Fine LJ, Green LW, Dougherty D. 2007. Study designs for effectiveness and translation research: identifying trade-offs. *Am. J. Prev. Med.* 33:139–54

72. Mohr DC, Schueller SM, Riley WT, Brown CH, Cuijpers P, et al. 2015. Trials of intervention principles: evaluation methods for evolving behavioral intervention technologies. *J. Med. Internet Res.* 17:e166

73. Monahan KC, Oesterle S, Rhew I, Hawkins JD. 2014. The relation between risk and protective factors for problem behaviors and depressive symptoms, antisocial behavior, and alcohol use in adolescents. *J. Commun. Psychol.* 42:621–38

74. Murphy SA, Lynch KG, Oslin D, McKay JR, TenHave T. 2007. Developing adaptive treatment strategies in substance abuse research. *Drug Alcohol Depend.* 88:S24–30

75. Murray DM. 1998. *Design and Analysis of Group-Randomized Trials*. Oxford, UK: Oxford Univ. Press

76. Murray DM, Lee Van Horn M, Hawkins JD, Arthur MW. 2006. Analysis strategies for a community trial to reduce adolescent ATOD use: a comparison of random coefficient and ANOVA/ANCOVA models. *Contemp. Clin. Trials* 27:188–206

77. Murray DM, Varnell SP, Blitstein JL. 2004. Design and analysis of group-randomized trials: a review of recent methodological developments. *Am. J. Public Health* 94:423–32

78. Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, et al. 2012. Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychol. Methods* 17:457–77

79. Natl. Res. Counc., Inst. Med. 2009. *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities*. Washington, DC: Natl. Acad. Press

80. Neumann MS, Sogolow ED. 2000. Replicating effective programs: HIV/AIDS prevention technology transfer. *AIDS Educ. Prev.: Off. Publ. Int. Soc. AIDS Educ.* 12:35–48

81. Nilsen P. 2015. Making sense of implementation theories, models, and frameworks. *Implement. Sci.* 10:53

82. Oesterle S, Hawkins JD, Fagan A, Abbott R, Catalano R. 2014. Variation in the sustained effects of the Communities that Care prevention system on adolescent smoking, delinquency, and violence. *Prev. Sci.* 15:138–45

83. Palinkas LA, Aarons GA, Horwitz S, Chamberlain P, Hurlburt M, Landsverk J. 2011. Mixed method designs in implementation research. *Adm. Policy Ment. Health Ment. Health Serv. Res.* 38:44–53

84. Patterson GR. 1974. Interventions for boys with conduct problems: multiple settings, treatments, and criteria. *J. Consult. Clin. Psychol.* 42:471–81

85. Pellegrini CA, Hoffman SA, Collins LM, Spring B. 2014. Optimization of remotely delivered intensive lifestyle treatment for obesity using the Multiphase Optimization Strategy: opt-IN study protocol. *Contemp. Clin. Trials* 38:251–59

86. Perl HI. 2011. Addicted to discovery: Does the quest for new knowledge hinder practice improvement? *Addict. Behav.* 11:590–96

87. Piantadosi S. 1997. *Clinical Trials: A Methodologic Perspective*. New York: Wiley

88. Poduska J, Kellam SG, Brown CH, Ford C, Windham A, et al. 2009. 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. *Implement. Sci.* 4:56

89. Poduska JM, Kellam SG, Wang W, Brown CH, Ialongo NS, Toyinbo P. 2008. 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 Alcohol Depend.* 95:S29–44

90. Powell BJ, McMillen JC, Proctor EK, Carpenter CR, Griffey RT, et al. 2012. A compilation of strategies for implementing clinical innovations in health and mental health. *Med. Care Res. Rev.* 69:123–57

91. Prior M, Elouafkaoui P, Elders A, Young L, Duncan EM, et al. 2014. 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. *Implement. Sci.* 9:50

92. Proctor E, Powell BJ, McMillen JC. 2013. Implementation strategies: recommendations for specifying and reporting. *Implement. Sci.* 8:139–50

93. Purtle J, Peters R, Brownson RC. 2016. A review of policy dissemination and implementation research funded by the National Institutes of Health, 2007–2014. *Implement. Sci.* 11:1

94. Rabin BA, Brownson EC. 2012. Developing the terminology for dissemination and implementation research. See Ref. 20, pp. 23–51

95. Raudenbush SW. 1997. Statistical analysis and optimal design for cluster randomized trials. *Psychol. Methods* 2:173–85

96. Reid JB, Taplin PS, Lorber R. 1981. A social interactional approach to the treatment of abusive families. In *Violent Behavior: Social Learning Approaches to Prediction, Management, and Treatment*, ed. RB Stuart, pp. 83–101. New York: Brunner/Mazel

97. Rogers EM. 1995. *Diffusion of Innovations*. New York: Free Press

98. Rosenbaum PR, Rubin DB. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55

99. Rubin DB. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* 66:688–701

100. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. 1996. Evidence based medicine: what it is and what it isn't. *BMJ* 312:71–72

101. Saldana L. 2014. The stages of implementation completion for evidence-based practice: protocol for a mixed methods study. *Implement. Sci.* 9:43

102. Shojania KG, Grimshaw JM. 2005. Evidence-based quality improvement: the state of the science. *Health Aff.* 24:138–50

103. Spoth R, Guyll M, Redmond C, Greenberg M, Feinberg M. 2011. Six-year sustainability of evidence-based intervention implementation quality by community-university partnerships: the PROSPER study. *Am. J. Commun. Psychol.* 48:412–25

104. Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. 2010. The use of propensity scores to assess the generalizability of results from randomized trials. *J. R. Stat. Soc.* 174(2):369–86

105. Szapocznik J, Duff JH, Schwartz SJ, Muir JA, Brown CH. 2015. Brief strategic family therapy treatment for behavior problem youth: theory, intervention, research, and implementation. In *Handbook of Family Therapy: The Science and Practice of Working with Families and Couples*, ed. T Sexton, J Lebow, pp. 286–304. Abingdon, UK: Routledge

106. Thistlethwaite DL, Campbell DT. 1960. Regression-discontinuity analysis: an alternative to the ex post facto experiment. *J. Educ. Psychol.* 51:309–17

107. Twardella D, Brenner H. 2007. Effects of practitioner education, practitioner payment and reimbursement of patients' drug costs on smoking cessation in primary care: a cluster randomised trial. *Tob. Control* 16:15–21

108. Valente TW, Palinkas LA, Czaja S, Chu KH, Brown CH. 2015. Social network analysis for program implementation. *PLOS ONE* 10:e0131712

109. Van Achterberg T, Schoonhoven L, Grol R. 2008. Nursing implementation science: how evidence-based nursing requires evidence-based implementation. *J. Nurs. Scholarsh.* 40:302–10

110. Weiss CH, Poncela-Casasnovas J, Glaser JI, Pah AR, Persell SD, et al. 2014. Adoption of a high-impact innovation in a homogeneous population. *Phys. Rev. X* 4:041008

111. Wyman PA, Henry D, Knoblauch S, Brown CH. 2015. Designs for testing group-based interventions with limited numbers of social units: the dynamic wait-listed and regression point displacement designs. *Prev. Sci.* 16:956–66

112. Wyrick DL, Rulison KL, Fearnow-Kenney M, Milroy JJ, Collins LM. 2014. Moving beyond the treatment package approach to developing behavioral interventions: addressing questions that arose during an application of the Multiphase Optimization Strategy (MOST). *Transl. Behav. Med.* 4:252–59