



Published in final edited form as:

*Annu Rev Public Health*. 2009 April 29; 30: 1–25. doi:10.1146/annurev.publhealth.031308.100223.

## Adaptive Designs for Randomized Trials in Public Health

C. Hendricks Brown<sup>1</sup>, Thomas R. Ten Have<sup>2</sup>, Booil Jo<sup>3</sup>, Getachew Dagne<sup>1</sup>, Peter A. Wyman<sup>4</sup>, Bengt Muthén<sup>5</sup>, and Robert D. Gibbons<sup>6</sup>

<sup>1</sup>Prevention Science and Methodology Group, Department of Epidemiology and Biostatistics, University of South Florida, Tampa, Florida, 33612; hbrown@health.usf.edu

<sup>2</sup>Department of Biostatistics, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, 19104

<sup>3</sup>Department of Psychiatry and Behavioral Science, Stanford University School of Medicine, Stanford, California, 94305-5795

<sup>4</sup>Department of Psychiatry, University of Rochester, Rochester, New York, 14642

<sup>5</sup>Graduate School of Education and Information Studies, University of California, Los Angeles, California, 90095-1521

<sup>6</sup>Center for Health Statistics, University of Illinois, Chicago, Illinois 60612

### Abstract

In this article, we present a discussion of two general ways in which the traditional randomized trial can be modified or adapted in response to the data being collected. We use the term adaptive design to refer to a trial in which characteristics of the study itself, such as the proportion assigned to active intervention versus control, change during the trial in response to data being collected. The term adaptive sequence of trials refers to a decision-making process that fundamentally informs the conceptualization and conduct of each new trial with the results of previous trials. Our discussion below investigates the utility of these two types of adaptations for public health evaluations.

Examples are provided to illustrate how adaptation can be used in practice. From these case studies, we discuss whether such evaluations can or should be analyzed as if they were formal randomized trials, and we discuss practical as well as ethical issues arising in the conduct of these new-generation trials.

### Keywords

multilevel trials; multistage trials; growth mixture models; encouragement designs; CACE modeling; principal stratification

## INTRODUCTION

Public health and medical care have long used the traditional randomized experiment, which pits two or more highly structured and fixed interventions against one another, as a fundamental scientific tool to differentiate successful interventions from those that are without merit or are harmful. Starting from James Lind's innovative 6-arm trial of 12 men who had scurvy (84), the conduct of trials underwent several transformations as they incorporated the revolutionary

ideas of experimentation developed by Fisher (47), the application to medical settings (66, 67), and the ethical protection of human subjects used in experiments (52,101). Journal editors and the scientific community expect that modern clinical treatment trials, as well as preventive or population-based trials in public health, will be conducted and analyzed to standards that maximize internal validity and their ability to produce rigorous causal inferences. These trial design protocols involve defining inclusion/exclusion criteria; assigning subjects or groups to interventions; exposing subjects to their assigned, distinct interventions; minimizing attrition; and conducting follow-up assessments that are blind to intervention condition. Traditional randomized trials also provide a fixed protocol, which specifies that intervention conditions remain constant across the study: a fixed number of subjects to be recruited, assigned to intervention, and then assessed the same way throughout the study. When a trial follows these protocols, it produces convincing causal inferences.

Despite the great successes that these traditional, well-conducted randomized trials have had in improving physical and mental health, such trials often have severe limitations. First, they are relatively poor in assessing the benefit from complex public health or medical interventions that account for individual preferences for or against certain interventions, differential adherence or attrition, or varying dosage or tailoring of an intervention to individual needs. Second, traditional trials virtually ignore the cumulative information being obtained in the trial that could potentially be used to redetermine sample size, adjust subject allocation, or modify the research question. Can such information that is being collected from a trial be used strategically to improve inferences or accelerate the research agenda in public health?

Researchers are starting to consider alternative trials as well as experimental studies that are not conducted as trials. We note three general ways that trials can adapt to information that is gathered and thus deviate from traditional trials. An adaptive design (sometimes called an adaptive trial or response adaptive design) uses the cumulative knowledge of current treatment successes and failures to change qualities of the ongoing trial. An adaptive sequence of trials draws on the results of existing studies to determine the next stage of evaluation research. An adaptive intervention modifies what an individual subject (or community for a group-based trial) receives as intervention in response to his or her preferences or initial response to an intervention. This article critically looks at the first two types of adaptive studies to evaluate public health interventions. Because of space considerations, this article includes only a very limited discussion of adaptive interventions.

Adaptive trials and sequences of trials are becoming more common in public health. These adaptations aim to address pressing research questions other than straightforward efficacy and effectiveness issues, although there is ongoing debate about the limitations that any randomized trial can provide in public health (128). In the following section, we present some examples of these alternative trials and include a typology of these alternatives, indicating when they may be appropriate and when they are not. We reference public health intervention trials whenever available, concentrating on that field's unique contributions to evaluating intervention impact on a defined population, on system-level or policy interventions, and on prevention. Such trials often use group or cluster randomization and require multilevel analyses of impact, in contrast with many clinical trials in the medical field (23). When good examples in public health are unavailable, we present examples from clinical medicine because many of the methods originated there.

## MOTIVATING EXAMPLE

For an example of an adaptive trial, we provide a brief introduction to two of the most contentious adaptive trials ever to be conducted. These trials both involved a novel intensive care treatment for neonates, the extracorporeal membrane oxygenation (ECMO) procedure,

which began to be used on infants whose lungs were unable to function properly by themselves (4). ECMO is a highly invasive procedure that oxygenates blood outside the body. Historical data on infants at one site suggested that ECMO was highly effective, with survival rates on conventional treatment without ECMO of 15%–20% and in similar patients treated with ECMO of 75% (91). However, because the scientific community heavily debated the appropriateness of this comparison based on historical controls, some neonatal units remained skeptical of its success and called for a randomized trial. Major ethical concerns were raised about a traditional randomized trial of ECMO versus conventional treatment, however, because the dramatic differences in response based on historical data strongly favored ECMO over no treatment. To reduce the number of babies who would likely die during a traditional clinical trial—where by definition half would receive the poorer treatment, which could well result in death—the first randomized trial (5,103) used a play-the-winner rule (134,140). Such an adaptive trial design tilts the allocation probabilities of intervention assignment from 50:50 to odds that favor the intervention that has so far performed better when one considers outcomes from all previous patients. The 1985 ECMO trial provides an example of how the play-the-winner rule can work. The first infant was randomly assigned to ECMO and survived. Because the first baby allocated to ECMO lived, the odds tilted 2:1 in favor of ECMO for the second child. Nevertheless, the second was randomly assigned to conventional treatment and subsequently died. On the basis of one failure under conventional treatment and one success on ECMO, the third subject's assignment had a 3:1 odds of being assigned to ECMO; this third infant was assigned to ECMO and survived. This tilting toward ECMO continued thereafter for the next 9 infants, who by random draw were all assigned to ECMO, and all survived. Thus the twelfth baby had odds of 12:1 of being assigned to ECMO. Even though the planned maximize trial size was 10, the study was terminated when the total sample size was 12. At the conclusion of the study, there was severe imbalance with only one subject assigned to conventional care who died (the second enrolled baby), and 11 subjects assigned to ECMO, all of whom lived.

### ECMO

extracorporeal membrane oxygenation

It would be hard to find fault with the primary motivation of the unusual assignment scheme in this trial. Their goal was to obtain clear causal inferences about the effect of ECMO versus conventional treatment in this population by denying the fewest number of subjects a potentially life-saving treatment. As results of a trial are continually tallied, should there not be a way to use these tallies to reduce the risk that the next subject will receive an intervention that results in an inferior outcome? This trial did lead to only one death, whereas a traditional trial might have led to more deaths. But there were major concerns about whether this trial achieved the first aim related to causal effects. Indeed, a standard two-sided Fisher exact test of these data reached 0.08 significance level, and many were troubled or unconvinced by any inference drawn on a single patient with conventional therapy in an unmasked trial. Just as important, serious ethical questions were raised about whether the study should have been conducted as a randomized trial in the first place and whether the Zelen randomized consent process (discussed below) was ethical (112,113,131). The Cochrane Collaboration, in its review of this trial, considers this trial as having a high potential risk for assignment bias (91).

The primary rationale that researchers use in proposing such adaptive designs and interventions states that there are some major shortcomings in the types of questions that can be answered by conventional randomized trials. However, a traditional randomized trial can accumulate enough data to make precise estimates of intervention effect by replicating the same process of recruiting and following subjects, assigning them in the same way to identical interventions across the entire study, and following strict analytic rules. Indeed, the rigid protocols used in

controlled clinical trials are what allow us to infer that any differences we find in the intervention groups' outcomes are due to the intervention itself and not to some other explanation. (The tradition in public health trials is typically not as rigid, particularly around the conduct of analyses.) This rigidity, however, comes with a price. Can trial designs be modified to increase recruitment, participation, or follow-up? Can the research question be further specified during the course of a trial? Despite the use of random assignment in these trials, it is not immediately clear whether any causal inferences can be drawn from such adaptations. Second, how can we best use the results of one completed trial to best inform the design of the next? Should there be two sequential trials, or is there a possibility of merging multiple trials into a single protocol that allows informed adaptation?

## WHAT WE MEAN BY ADAPTATION

The word adaptation has been used in many different and often in conflicting ways when applied to randomized trials, a fact that becomes apparent in a routine literature search of this term. Thus it is important to clarify what we mean by this general term and provide a useful typology that incorporates what is being adapted and where it is applied. Building on the work of Dragalin (40), we conceptualize three distinct types of adaptation. Some of these adaptations affect specific elements of the study design of an ongoing trial, some relate to the next trial, and still other adaptations relate to the intervention or to its delivery. All these usages share a critical trait that for us defines adaptation: As data are collected, they are used informatively to make important decisions pertaining to a trial. These decisions are anticipated, and the alternative choices are enumerated in advance. Some of these change points are sudden, such as stopping a trial because of clear evidence of superiority, whereas others involve more or less continuous change throughout the trial. They can be done for reasons of design efficiency, for example, saving time by terminating a study early or recomputing a required sample size, or for ethical reasons, by limiting harmful outcomes or providing treatments that are in concert with patient preferences or likely benefit. Care is taken in these adaptive protocols by attempting to protect the statistical properties of the inferential decisions regardless of which decision is forced by the data.

The three meanings we offer depend on the use of these decision points. Specifically, these informed decision points may affect (a) the design of a new trial, (b) the conduct of an ongoing trial, and/or (c) the intervention experience of a study participant. To distinguish these types of adaptations, we refer to them in the adjective form as adaptive sequencing (of trials), adaptive designs (of trials), and adaptive interventions, providing further specification with adjectives as needed. Our description begins with adaptive designs.

## ADAPTIVE DESIGNS

Adaptation in the context of an ongoing trial refers to planned modification of characteristics of the trial itself based on information from the data already accumulated. Thus the trial design for future subjects is adapted using this information. A number of different adaptive designs are listed in this section, and examples relevant to public health are given below.

One example of such an adaptive trial is illustrated in Figure 1, in which subjects begin the trial with equal random assignment to Arm 1 or Arm 2. After a planned interim analysis, there may be sufficient evidence that one intervention is already found to be superior to another or that it is virtually impossible that a comparative test will find significance by the projected end of the trial. In these cases, it may make sense to end the trial early. Another alternative, shown in the bottom part of the figure, finds that while the results are not yet conclusive, the data may support Arm 2, and therefore a play-the-winner rule could adjust the probabilities of assignment to favor Arm 2.

In Table 1, we present a range of adaptive designs, arranged by the characteristic of the design that is adaptive; which data are used in these decisions; and benefits, requirements, and cautions when using these methods. A key reference is provided with each entry. Below we describe each of these areas in more detail.

### Early Stopping

A large clinical trial literature documents what are known as sequential trials, for which detailed protocols, written before the start of the study, allow for early stopping or other changes based on findings to date. Early stopping in sequential designs is based on planned interim analyses (135). Each interim analysis accounts for a piece of the Type I error so that a highly significant finding is required to stop the trial at the first planned interim analysis and a less significant finding required to stop the trial based on succeeding interim analyses. If no interim result is sufficiently definitive to stop the trial, some of the Type I error rate is already used up, so the final analysis at the intended end of the trial has to pay a modest penalty for these early looks. Bayesian methods are also available that update prior beliefs to form posterior distributions about the interventions' relative effectiveness on the basis of all available data. Bayesian methods are also used to assess what the results are expected to be for the full trial by way of the predictive distribution (10–12,69,119,143).

Although these sequential designs are common in clinical medicine trials, they are very uncommon in public health trials for three reasons. First, interim analyses require subject recruitment to be sequential; therefore, this excludes group-based cohort trials, such as school-based randomized designs, where whole cohorts are enrolled at the same time. Second, public health often concentrates on prevention rather than treatment, and tradition has it that the cost of subjecting participants to a less beneficial preventive intervention or control condition is not nearly as harmful as exposing more people to a worse treatment. Third, many public health trials take seriously secondary aims, such as testing for moderating effects across subgroups. Stopping a trial early when the main effect is significant would typically not allow sufficient statistical power to test for interactions.

### Optimal Dose

Finding an optimal intervention dose is another common reason for conducting adaptive trials, especially for pharmacotherapies (41,42,54,55,79,102,124). Typically, one begins with subjects allocated to one of several dosages or placebo, then an analysis of the dose-response relationship is made, from which new dosages are set so that the next set of subjects provides optimal information on a key effective dose. Direct usage of these designs in public health trials is also rare. One reason is that such strategies require that responses be measured shortly after intervention to inform the design of the next cohort. One could consider whether optimal dose designs would be useful to tackle some of the challenging public health problems, for example, to examine the dosage of aspirin that produces the best outcome on cardiovascular disease. All the large aspirin prevention trials conducted to date have tested placebo against a single dose (44,109). These doses in these preventive trials, however, have ranged from 75 mg per day to 500 mg per day. The only practical way that an optimal dose trial could work is if a strong, early surrogate for heart disease were available. Similar barriers exist when determining appropriate dosages of vitamin supplementation.

Opportunities for some dosage trials with psychosocial or public health policy interventions are possible as well. Two examples are Internet-based randomized trials for smoking cessation and prevention of depression, where dosage can refer to the intensity of supportive communications (92). The large number of randomized subjects in these trials, the limited time frame for success (i.e., 1–3 month outcomes), and the ease in varying the level of emailed reinforcing messages make it very feasible to adapt the dosing strategy. However, many public

health interventions have built-in limitations to the choices of dosage and are therefore not appropriate to apply these methods. For example, one trial tested daily versus weekly supplementation for pregnant mothers, the only two reasonable alternatives to dosage (45).

## Sample Size

Another common reason for conducting adaptive designs in clinical trials is the redetermination of the trial's sample size using the data at hand. The most common (but not often the best) way to determine sample size is to decide on an effect size for the intervention that if true would achieve 0.80 power. A second approach to determining sample size—or sample sizes across multiple levels for group-based trials—is to specify the required power for the trial and use a formula that relates power to sample size, the change in the mean level of the response variable produced by the intervention, and the variability in the measure. The mean response level of the intervention is chosen as a clinically meaningful value. If we have a good estimate of the variance, then the required sample size for the study can be computed. Often, however, we may have limited information about the underlying variance, and if one were to guess wrong, the calculated sample size would be very inaccurate. It therefore may make sense to use data in the existing study to reassess the variance, thereby modifying the required sample size.

A two-stage adaptive approach would allow the first-stage data to be used to estimate either the effect size or, alternatively, the variances, thereby recomputing an appropriate sample size (116). Care must be taken, however, to maintain the type I error rate, otherwise spurious findings can occur. The problem of inflated type I errors can be avoided if the sample size is based only on the variance rather than the estimated effect size. Other statistical issues are discussed in References 60,88, and 108.

To date, few public health trials have recomputed sample size during the trial, despite the potential benefit of doing so. One particularly promising set of opportunities comes from group-based randomized designs in which misspecification of the intraclass correlation can lead to vastly underpowered or overpowered designs (38,39,80,95). Lake et al. (80) provide a method for reestimation of sample size for such studies. These ideas can also be used in trials that involve repeated measures designs, such as those evaluating how an intervention changes the slope of growth trajectories (98). In such studies, the hypotheses are often well specified from the theory of the intervention itself, but because power depends on so many variances and other secondary parameters, each of which is difficult to assess, it would be useful to revise the size of the study after some data are produced. When a trial has both clustering and longitudinal or repeated measures, statistical power is an even more complex function of the numbers of subjects, clusters, and follow-up times, as well as intervention impact on means and variances, but programs are available to evaluate power or recomputed sample size (13,111). With these tools now available, it would be appropriate to consider two-stage designs that modify their recruitment of the numbers of subjects or the number of clusters on the basis of improved estimates of these quantities.

Another reason to modify the size of a trial on the basis of collected data is because there may be evidence of an interaction involving the intervention. Ayanlowo & Redden proposed a two-stage design to address this situation (2). In this type of design, the first stage uses a standard random allocation without regard to covariates and assesses whether there is evidence of any interaction between the covariates and baseline. If there is, the second stage takes account of covariates through stratified random sampling.

## Enrollment Designs

We use the term enrollment design to refer to a trial during which individuals are randomized to different protocols for enrolling in the trial itself, rather than to different intervention

conditions. One of the major challenges in delivering interventions in public health occurs because a large portion of subjects may be disinclined to participate; enrollment trials provide a way to test different strategies for improving participation rates. In addition to low enrollment rates, many public health trials show differential rates of enrollment across sociodemographic groups as well as risk status. Consequently, there could be selection effects that limit the generalizability of a trial. Indeed, the low and declining enrollments of minorities in cancer trials has raised considerable concerns about the effects of treatments on these groups (96). Investigators do use certain strategies to increase the proportion of subjects who participate in a trial, and we briefly review these here. One approach to modifying the enrollment procedure is to randomize subjects to an intervention condition, then ask for consent to be treated with that specific intervention. In trials with one new intervention and a control condition, those randomized to the intervention can receive the control condition if they refuse; those randomized to the control condition are not offered the intervention. This type of randomized consent (or randomized before consent) design was proposed by Zelen (140–142) and used in the early ECMO study cited above (5). The original purpose of this design was to overcome physician reluctance to asking patients to agree to be randomized. A number of important ethical issues need to be addressed when using these types of designs, particularly issues around informed consent and beneficence. And mostly because of these concerns, such randomized consent designs have rarely been used in clinical treatment trials. A few prevention, rather than treatment, trials did use such designs because the issue of withholding a potentially beneficial program from nondiseased subjects is of less urgency than it is with treatment trials. Indeed, Braver & Smith (17) proposed a “combined modified” trial design with arms involving both traditional randomization after consent and randomization before consent.

Although randomized consent designs do not use any subject data to modify enrollment procedures, some enrollment designs do make use of such data and can therefore be considered adaptive. One example is the use of motivational interviewing (46) to increase enrollment into a wait-listed trial, that is, one in which consented subjects receive the same intervention either immediately or later. The purpose of this type of study would be to test whether the invitation process itself—in this case, motivational interviewing—increases enrollment and whether this effect on enrollment diminishes if there is a waiting period. Designs of this type are being carried out by the Arizona State Prevention Research Center, which is testing a parenting program for preventing mental and behavioral problems in children of divorce through the court system (136,137).

Enrollment in randomized trials is shaped for two general reasons, and these reasons have opposing objectives. First, investigators want to tune the trial so that it has higher statistical power by selecting certain subjects who are more likely to consent to the trial itself, to adhere to the assigned intervention, to benefit from the intervention, or to agree not to drop out of the trial. If a researcher can increase the rates of these factors, the design will have increased statistical power. In this approach, individual-level data are used to model these issues separately, and on this basis, subjects are selected for inclusion in the trial. We discuss a number of these approaches below. The second alternative reason for shaping enrollment in randomized trials is to expand the population for purposes of maximizing potential benefit of an intervention at the population level as it is adopted by communities. This approach is associated most closely with a public health model, such as that embodied by RE-AIM (56,89) and a population preventive effect (21,58). A comparison of the relative strengths and weaknesses of each approach is found in Reference 20.

### **Mediational Selection Designs**

Pillow et al. (106) planned a randomized trial that was restricted to those subjects who had low scores at baseline and, therefore, had higher potential improvement on hypothesized mediators

of a family-based intervention. This design sought maximal overlap between the population subset being tested and the subset anticipated to experience the highest benefit from this intervention.

### Adherence Selection Designs

Similar to the last example, one could build a statistical model that would predict participation or adherence to one or both intervention conditions. The principal stratification approach to causal inference (1,49,50,53,70–73,74–76) focuses primarily on this group of adherers or compliers as the group that will provide most of the causal information about the intervention. An adherence selection design would then exclude those subjects who are not likely to participate as assigned. It is possible to conduct a trial design in two stages: The first would be used to develop a model of adherence, which would then be used to form new inclusion/exclusion criteria.

### Response Selection Designs

A related approach is to use covariate information to predict whether a subject is likely to respond on the control (or placebo) condition; if so, then such individuals will not contribute much to overall impact. In some placebo randomized pharmaceutical trials, the proportion of placebo responders is often quite high, so higher effect sizes could be obtained if one could exclude only those on placebo. Despite the question of whether such selection is appropriate from a scientific standpoint, it appears difficult to obtain good baseline covariate predictors of placebo response; there is a better chance to assess treatment responders, however, at least for antidepressant medications (97).

### Run-In Designs

In the previous examples, we relied only on baseline data for all subjects to determine whether they should be enrolled in the trial. However, opportunities also exist to use repeated measures prior to random assignment to determine who should be eligible for random assignment. This idea was used in the Physician's Health Study to exclude subjects who were not likely to adhere to their respectively assigned interventions. Fully one-third of the subjects were excluded from randomization by using an 18-week run-in to examine their adherence to beta-carotene placebo/open-label aspirin (104). The use of run-ins for maximizing adherence has been looked at from a methodologic point of view, as well (81,115).

Investigators have also used run-in designs to identify placebo responders more efficiently than is possible with baseline covariates alone. Although this procedure has been used in a number of studies, it has evoked controversy (9,83).

Another use of a run-in design is to reduce attrition by selecting a sample that is nearly immune to drop-out (127). Both of these run-in trials involved individual-level randomization, but there is also a potential benefit for their use in group-based randomized trials. For example, to test whether a classroom behavior-management program is effective, one could take baseline measures on all classrooms then select only those classrooms that have high levels of aggression as eligible for inclusion in a randomized trial. By limiting the study to classes that have poor behavior management at the beginning and by reducing the variability across classrooms, the power of this design should increase (D. Hedeker, personal communication).

### Encouragement Designs

A growing number of so-called randomized encouragement designs have been used in public health to encourage behaviors related to participating or adhering to an intervention (43). Each arm in the trial exposes subjects to the same (or similar) intervention, so the fundamental



difference between the arms are these different encouragements. Encouragements can take the form of reminders (36,87), lotteries, or incentives (3,18,129); incorporation of preferences (122); additional supportive training; technical assistance or coaching (37); case managers; or sophisticated social persuasion techniques (32). The encouragements can be directed toward individuals (129) and families (138), physicians (15), other health care providers, teachers (37), schools (105), villages, or county-level governmental agencies (29). These designs address two main scientific questions. First, they test whether participation in or adherence to an assigned intervention can be modified using incentives, persuasion, or other types of motivators. If these techniques are cost-efficient, they could lead to a larger impact within a community. Second, these designs can also be used to test the causal effect of an intervention when there is less than perfect adherence. Although a full discussion of these causal inference techniques is beyond the scope of this article, we can make an intuitive connection between encouragement designs and causal impact. If an encouragement strategy increases overall adherence or participation in an intervention, then we would expect that the beneficial response in that arm relative to that with low encouragement would depend on the true improvement in the program and the proportion of the sample whose adherence behavior was dependent on the level of encouragement they received. Statistical methods for complier average causal effects (CACE) and related instrumental variable approaches were developed to account for self-selection factors that would otherwise be hidden within standard intent-to-treat analyses (1, 14,48–51,57,68,71,76,99,118,120,121).

### Adaptive Allocation Assignments

Response adaptive designs (RAD) use the success or failure results on all previous subjects in each intervention condition to modify the allocation probabilities, steering successive subjects toward the perceived, preferable treatment or alternatively enforcing a restriction on the number of subjects who have poor outcomes in each arm. These procedures include the continuous adapting play-the-winner rule, which was described in the first ECMO trial above (5). A second example includes the two-stage adaptive design that O'Rourke et al. (103) used in a follow-up trial of ECMO. This design guaranteed no more than a fixed number of deaths could occur in either condition. In this case, they designed the first phase of the trial using equal allocation probabilities. This would continue until four patients died on one of the intervention arms. Then the study would transition into the second nonrandomized phase of the study during which patients were assigned only to the intervention with lower mortality. When the first randomized phase of the trial ended, the nominal significance level was 0.087. The research team continued enrolling patients but assigned them all to a better treatment—ECMO—thus ending any random assignment completely. Although the physicians were not explicitly told that randomization had stopped, it soon became apparent because masking was not possible. The stopping rule for this second stage was to continue enrolling in ECMO until the twenty-eighth survivor or fourth death occurred. The trial stopped with 20 additional patients enrolled; of these only one died (131,132). The nominal significance level when data from both phases were combined was 0.01.

This trial generated a great deal of discussion, ranging from criticisms that the trial itself was unnecessary to begin with and caused four unnecessary deaths (112) to questions about whether randomization should have been stopped when it did (6). Eventually, a large traditional randomized trial was conducted in the United Kingdom that showed convincing one-year and long-term benefits on both mortality and severe disability (8,126).

RADs of the type used in the ECMO trials have been criticized because they could introduce a major bias if a time drift occurs in the types of patients recruited or if one of the interventions changes across time. Indeed, this design severely compounds the effects of a drift toward lower-risk patients when using play-the-winner rules. This compounding can occur even when there

is no difference in outcomes across the interventions because the lower-risk patients would likely survive and thereby quickly increase the allocation probability for one of the interventions (30). Indeed, both of these ECMO trials ended with a complete confounding of time of entry with intervention condition. A recent proposal for a group-sequential RAD, which forces every block of time to have random assignment to both intervention conditions, is likely to resolve this problem from the design point of view (77), and we can recommend that the design be used when ethical and contextual factors permit.

To our knowledge, RADs have not been used for trials other than those in medicine, where they are still quite rare, nor for any group- or place-based randomized trials (16,23,95) that are common in public health. We believe they could be used in other settings. Consider, for example, a state that wants to implement a school-based suicide-prevention program that has yet to receive any rigorous evaluation. Officials may expect that this program would reduce the number of youth dying from self-harm and therefore would feel compelled to train as many schools in their state as possible. There are legitimate reasons to question whether some of these programs may even cause harm (61–64). But a statewide implementation in schools, if done by randomly assigning the times that schools obtain training, could provide valuable scientific information about the program's effectiveness on completed suicides that, compared with other health outcomes, occur so rarely that randomized trials are difficult to conduct (24). A large group level RAD could be conducted by using two levels of randomization. First, counties could be segmented to form separate sets of cohorts consisting of all schools in a group of counties, and counties within each cohort would be randomized to receive the intervention or not (or to receive the intervention later in a cohort wait-listed design). The first cohort would start with equal numbers of schools assigned to the trained and nontrained condition, assessing all completed and hospitalized attempted suicides within all those counties in that cohort. In successive cohorts, the subtrials would then adjust the allocation probabilities in light of the cumulative findings. The analysis would incorporate cohorts as a stratified or blocking factor in the analysis. If the findings determine that the training does reduce completed and attempted suicides, then the formal trial would be stopped and training would occur for all schools; alternatively, if the findings indicate null or harmful effects, all training would stop.

### **Other Adaptive Intervention Assignments Based on Covariates**

Wei (133) provides a useful method for balancing a study design throughout a study across many different covariates. This technique assigns new enrollees in such a way as to minimize the imbalance across the least balanced variable. This method could be put to good use in some public health trials, during which many potential confounders exist but cannot all be easily controlled. More recent approaches to dealing with imbalances across high dimensions of covariates tend to involve the use of propensity scores, which are model-based estimates of probabilities of assignment to intervention based on all covariates (110). Hahn and colleagues (65) consider the use of propensity scores in an adaptive way. They form a propensity model on the first block of data and use that to make adjustments.

### **Follow-Up Designs**

As a final type of adaptive trial design, we consider modifications to the follow-up design of a randomized trial based on data that have already been collected. An efficient way to evaluate the impact of the intervention on a diagnostic condition that is expensive to assess is to use a two-stage design for follow-up. The first-stage data are obtained on all participants; this stage includes an inexpensive screening measure. All screen positives and a random sample of screen negatives can then be assessed on the more expensive diagnostic assessment. Standard missing data techniques will then allow investigators to compare the rates of these diagnoses by intervention condition (23). They can also conduct stratified follow-up using not only the

outcome data but also the pattern of missing data. Such strategies can provide additional protection against nonignorably missing data (22).

## ADAPTIVE SEQUENCING OF TRIALS

Ideally, the research community carries out trials that either complement one another, such as replication trials, or build on the results of previous trials to improve our knowledge of what works, under which conditions, and how to move this knowledge into practice. In either case, trials should be conducted to provide empirical evidence for sound decision making. Thus we would always expect the results of completed trials to inform the next stage of scientific evaluation, whether to replicate this trial in some form, whether to follow up the same subjects to evaluate longer-term impact, whether to focus on a new research question that leads to a different research design, or whether to end this line of research because questions are sufficiently answered or trials in this area are no longer ethical or fruitful. For all these decision points, a type of adaptive sequencing occurs after a trial concludes and thus affects the construction of future trials. In essence, all sequences of trials should have an adaptive quality because they depend on previous trial results. The following section is an illustration of a traditional model for trial sequencing; we then present some more innovative and lesser known, adaptive sequences.

### Traditional Sequence of Trials for Developing, Testing, and Implementing an Evidence-Based Program

We now consider the stages of research that are necessary from initial program development and testing to full implementation of a successful program. The Institute of Medicine's (IOM) modernized prevention-intervention research cycle (90) provides a set of scientific steps, often involving randomized trials, that guides the development and testing of an intervention for efficacy and effectiveness followed by wide-scale community-level adoption of this program. In this research cycle, beneficial findings on small efficacy trials of an intervention would likely lead researchers to conduct trials that move toward real-world evaluation, or effectiveness. If the finding is strong enough, the intervention may be identified as an evidence-based program, in which case the emphasis may switch toward testing alternative ways to convince communities to adopt and deliver this intervention with fidelity. It would then be reasonable to conduct an implementation trial, that is, a trial that tests different strategies for implementation. The arms of such a trial consist of different ways in which this program engages communities and facilitates their use. Alternatively, if there are null or negative findings, researchers may be forced to cycle back to earlier stages.

One example of this typical sequencing of trials is provided by the testing and implementation of a successful program for supporting foster children who have significant behavioral and mental health problems using the multidimensional treatment foster care (MTFC) program (28). This model was tested in several randomized trials (29) and, as a result of these efficacy and effectiveness trials, has now been recognized as an evidence-based program. But interventions that are found to be successful in randomized trials do not necessarily result in program implementation. Indeed, when all counties in California were offered training in this program, only 10%—the so-called early adopting counties—implemented this program. To address the question of whether nonearly adopting counties could achieve full implementation through technical support, the research team began an implementation trial, during which all 40 eligible counties in California were randomly assigned to two alternative strategies for implementing MTFC (29). This is one example of the typical sequence of trials that began with an efficacy trial, moved into an effectiveness trial, and then was followed by this current implementation trial.

## THE PILOT-TRIAL-TO-FULL-TRIAL DILEMMA: WHEN ADAPTING CAN LEAD TO PROBLEMS

Given the large expense incurred by randomized trials, it only makes sense to invest the limited scientific capital we do have in interventions that are promising. But there is no easy way to identify which interventions are promising. One common approach is to conduct a small pilot randomized trial and to use an estimate from this small pilot trial, such as an effect size, to determine the sample size required for the full trial. This full trial design is then presented to scientific reviewers who then determine whether this particular intervention deserves funding. The National Institutes of Health (NIH) has specific funding mechanisms, such as the R34 (Exploratory Research Grant Award) and the R21/R33 (Phased Innovation Grant Award), that support the collection of such pilot trial data to determine if the intervention is worthy of the traditional R01 funding for a large trial. The operating characteristics of this pilot-to-full-trial design and decision process, however, have raised some serious concerns (78). Too many interventions that are truly effective will show up as having low effect sizes from the pilot study (and thus rejected for full funding), and for those interventions that are selected based on having a large estimated effect size, the resulting sample size will often be too small, thus leading to an underpowered study. Thus this underpowering could greatly reduce the chance that a truly effective intervention will be found in the full trial to have significant benefit.

The main problem is that the effect size estimate (or any other measures of intervention impact) that is obtained from the small pilot trial has too much variability to be used as a screen to sort out potentially strong versus weak interventions and determine sample size. These problems can be quite severe. If, for example, the intervention has a true effect size of 0.5, a one-sided test at 80% power would require 100 subjects. If a pilot trial of 50 subjects was used to determine sample size, the true power for the larger trial, whose sample size is determined by the pilot study, is reduced to 0.60 (78).

How might these problems be overcome? First, if there was no attempt to adapt the full trial using pilot trial results, these types of biases would not occur. Second, funding decisions and evaluations by reviewers should be aware of the limited precision of effect size obtained from small studies. Third, researchers should view their own effect size estimates with some caution, as well, and should build in larger studies if the effect size estimate looks higher than one would reasonably expect.

### Adaptive Sequencing of Trials by Merging Multiple Stages of Research

The preventive intervention research sequencing from small to larger trials illustrated above has a parallel in pharmacologic research, in which the familiar sequencing of the traditional phase I–III trials allows investigators to identify biologically active drugs, dosages of biological significance, basic safety information, and initial screening of promising from unpromising drugs, followed by careful evaluation of effectiveness and safety of a drug delivered in a standardized fashion (107). See Table 2. One adaptive approach that deviates from this traditional piecemeal sequence is to combine two steps into one smooth approach using an adaptive procedure. For example, phase I/II adaptive trials allow researchers to jointly answer questions of toxicity, which is typically a phase I question, and to determine effective dosage levels, which typically occurs in phase II (7,31,41,54,59,123,124,144). The goal of this type of integrated trial is to identify quickly drugs that are not safe or biologically responsive and reduce the size and duration of the overall study. Other similar research studies also combine phase II and phase III trials (31,54,69,85,125,130).

Public health research has presented two related criticisms of the traditional preventive-intervention cycle approach, which breaks the research agenda into the predictable sequence

of trials described above. These criticisms have in turn led to some adaptive sequences of trials. First, the traditional approach often delays getting feedback about community acceptance of the intervention until a program has been found to be efficacious, at which point the program and its delivery system may be so established that it is hard to integrate community concerns. As a consequence, communities may be reluctant to adopt it, or individuals may not participate, thereby reducing the impact at a population level (56,89). The second related criticism of the traditional preventive-intervention cycle is the time it takes investigators to develop and test such programs. For a whole list of social and behavioral problems, such as substance abuse prevention, violence and delinquency prevention, injury prevention, suicide prevention, as well as lists of medical conditions, such as obesity, diabetes, cancer, and hypertension, advocates, community leaders, and policy makers will not wait patiently for the slow progress of science. Instead, the need to do something is so great that there is often a push for wide-scale implementation of a program that has not been subjected to rigorous scientific evaluation. We propose that these roll-out situations provide valuable opportunities in public health to adapt the sequence of trials. We briefly discuss an example used in preventing youth suicide, an area in which no interventions have yet demonstrated clear benefit at the population level.

### Roll-Out Randomized Trials

A roll-out trial allows investigators to test an intervention for effectiveness after a decision has been made to implement the intervention system wide. Although the scientific tradition usually conducts small efficacy and larger effectiveness trials before deciding to engage in wide-scale implementation, a number of situations have occurred where the decision to implement comes without any formal testing. Public health interventions are often mounted across broad communities, and this process provides a good opportunity to evaluate an intervention as it is rolled out. One recent example involves an untested suicide-prevention program. Gatekeeper programs train broad segments of the population as a means to increase referrals of actively suicidal persons in that population to mental health services. Investigators hope that widespread teaching of warning signs of suicidality, training in how to ask directly if a person is suicidal, and learning the how and where of referring suicidal individuals will result in improved treatment and ultimately reduce the number of suicides. In Georgia, the state legislature determined that gatekeeper training should be implemented in some secondary schools. One large school district chose to implement QPR, a well-known gatekeeper training program. This district began by implementing QPR in two schools that had recently experienced a youth suicide but were persuaded by research advisors that a full randomized trial could be constructed to test effectiveness by randomly choosing when each of the 32 schools' training would begin. This dynamic wait-listed design (25) satisfied not only the researchers' needs but also those of the district. In such a design, a randomly selected set of schools received QPR training first, followed by other sets of randomly assigned schools until all schools were trained. Randomization of the training of schools by time interval allowed a rigorous test of QPR, which avoided potential confounding by school readiness and time-dependent seasonality effects of suicide, as well as media coverage of high-profile suicides. This design provided more statistical power than that available in the traditional wait-listed design (25). The school district was satisfied with this design as well. By randomizing the intervention units (schools) across time, the district was assured that all schools would receive training by the end of the study, which would not be true with a traditional control group. Furthermore, the randomization of small numbers of schools to receive training during a time interval greatly reduced the logistical demands of training. Intent-to-treat analyses for multilevel designs (23) demonstrated that knowledge and attitudes of the school staff were dramatically increased, with limited change in self-reported referral behavior (139). Trials known as stepped wedge randomized trials (19) have used designs similar to the dynamic wait-listed design described here.

## Withdrawal and Roll-Back Randomized Trials

Researchers clearly prefer to use randomized trials to test for efficacy of a new drug compared with placebo or a standard drug. It is much rarer for scientists to test whether a drug that is being taken continues to provide ongoing benefit or whether its usefulness has ended. Archie Cochrane, whose inspiration for summarizing the trial literature led to the development of the Cochrane Collaboration, was known for seeking answers to such questions. Despite the potential benefit of more withdrawal trials, where subjects have all been taking a drug, and half are randomly assigned to withdraw from the drug, the few withdrawal trials in the literature generally occur when a drug is causing adverse events. A recent withdrawal trial of an antiepileptic drug was done and found that withdrawal led to poorer outcomes (86).

Public health trials could benefit from what can be termed roll-back trials, which would have some similarity to withdrawal trials. In times of tight budgets, policy makers often make critical decisions to cut back or eliminate community-wide public health programs. Even with such retrenchment, an opportunity exists to conduct a careful evaluation of an intervention, provided there is random assignment to which units are the first to give up. This type of trial would be a roll-back trial.

## Rerandomization to New Interventions

The typical approach to evaluating a new intervention is to begin with a new sample of subjects. It is possible, however, to conduct a new trial using subjects who were already part of an earlier trial. An example of this is the follow-up of the Carolina Abecedarian Project (27), which initially randomized high-risk infants around four months of age to an early center-based educational/social development program or to the control condition. Years later, when the children were close to entering kindergarten, they were rerandomized to a home-based learning program or control. Thus this study, which began as a two-arm trial, evolved into a two-by-two factorial design. The goal of the study was to examine the differential effects of early and later intervention, as well as the cumulative effects of the two interventions.

We note that there are many examples of individual-level rerandomizations based on failure to respond to treatment; however, because these designs fit more closely into adaptive interventions, we provide only a brief review here. Within-subject sequential randomization designs have been implemented to evaluate adaptive treatment strategies involving sequences of treatment decisions for individual patients, where decisions incorporate time-varying information on treatment response, side effects, and patient preference (93). Under these designs, each participant is randomized multiple times at each decision point in the adaptive treatment strategy (33,94). Such designs require a sequence of very specific rules and adherence to these rules for when and how treatments should change. Each rule leads to the randomization of participants to several treatment options. Randomizing a given patient at each rule-based decision point allows comparisons of treatment options at each decision point as well as comparisons of strategies or pathways encompassing all or subsets of decision points. Standard baseline randomization studies during which only the initial treatment assignment is randomized but treatment rules are administered subsequently (26) require the assumption of no unmeasured confounding when evaluating these post-baseline rules. In contrast, the sequential randomized designs do not require such an assumption because of the randomization of treatment options for each rule. The rules are based on ongoing information from the patient under treatment such as patient response status to current and past treatment, side effects, adherence, and preference toward treatment. The rules consist of tailoring or effect modification variables that stratify the sample at any point in time corresponding to a treatment decision. Within each stratum, the patients are then randomized to a set of treatment options. Lavori & Dawson (82) called such groupings equipose stratification. The tailoring or stratification is done to create homogeneous groups with similar disease status and equipose

with respect to the treatment options offered. Outcome responses to the randomly assigned treatments may then be used to form new strata or tailoring groups within which participants are randomized with respect to the next set of treatment options determined by the tailoring variables.

### **Multiphase Optimization Strategy Trials**

Collins and colleagues (34,35,100) have proposed a method for efficiently screening effective intervention components in large trials such as one that is likely to run using the Internet. With even a small number of possible components to include in a final intervention, an exponentially increasing number of combinations are possible. To quickly assess their merit while allowing for interactions with other components, these authors suggest using a fractional factorial design. These designs select a fraction of all possible combinations of intervention conditions, balanced across lower-order effects, and randomly assign subjects to these combinations. To decompose the individual and combined effects, researchers use analytic models that assume that high-order interactions are negligible. This screening for effective components can then be followed by a refining phase, during which dosage and other factors are examined.

### **Cumulative Adaptive Trials**

Another approach to combining trials adaptively is to continue improving an intervention through repeated trials while using the cumulative data to conduct an overall evaluation of the intervention, even as it is refined. We provide an example of recent work on evaluating a second youth suicide prevention program. In evaluating the QPR gatekeeper training program described above (25,139), researchers found no indication of its success in high schools. These limited findings in high school from the previous study on suicide prevention persuaded researchers to move down the preventive intervention research cycle to begin evaluations on a different intervention. The Sources of Strength program uses peer leaders to modify a school's norms about disclosing to trusted adults that a friend is suicidal. Like QPR, it has relatively limited empirical evidence behind it, although it has been identified as a potentially promising program. This intervention is currently undergoing testing and refinement using smaller randomized trials using eight or fewer schools within a state, whose times of intervention initiation were randomly assigned. We use the term cumulative adaptive trial to describe the strategic plan behind these sets of smaller trials. The first goal of these sequences of trials is to test in individual small trials specific parts of the theory of the Sources of Strength program's impact on hypothesized mediators and early proximal outcomes. For example, in the first such trial, involving six schools, training of peer leaders significantly changed students' willingness to disclose their friend's suicidality to trusted adults. Each of the new trials is testing a small portion of the model, and the results are used to improve the program's quality (see Figure 3). The second goal in this sequence of trials is to test the Sources of Strength's effects on more distal measures, including referrals for suicidality, attempts, and eventually completed suicides. Because all these events are relatively rare, each of the smaller studies is far underpowered to conduct an experiment. Their combined data from a number of these school/time randomized experiments do, however, provide increasing statistical power to detect impact on these comparatively low base rate outcomes (24).

This cumulative sequence of trials attempts both to improve quality over time and to evaluate intervention impact. One could legitimately question whether a statistical analysis to combine such data, such as that used in meta-analysis, would lead to appropriate inferences when the intervention may be changing over time. However, the statistical properties are clear that as long as the occurrence of these changes does not lead to a decrease in effectiveness, type I errors can be controlled.

## CONCLUSIONS

This review has discussed a wide range of adaptive trials that can be of use in public health. Although we have briefly discussed the relative merits and weaknesses of these trials, researchers will still need to consider carefully their value for their own work. We have limited our discussion here to randomized trials. However, investigators can find opportunities to combine trials with other kinds of research designs. One set of models now beginning to be used in public health is the product development model, which has been borrowed from business. This model involves a strategic plan to develop a successful program and bring it into practice by involving the consumers in program development at all stages. It contrasts with the traditional scientific approach that assumes demonstrating benefit early on is the priority. We recommend integrative approaches that combine much of the qualitative data that come from these approaches to trial data (114,117).

## Acknowledgments

We thank members of the Prevention Science and Methodology Group for their helpful discussion in the development of this paper. We also gratefully acknowledge the insightful discussions of Drs. Sheppard Kellam, Helena Kraemer, Phil Lavori, Don Hedeker, and Liz Stuart, all of whom provided important insights for this paper. We thank Ms. Terri Singer for conducting library searches for this review. Finally, we acknowledge the support of the National Institute of Mental Health and the National Institute on Drug Abuse under grant numbers R01-MH40859, R01MH076158, R01MH076158, P30-DA023920, P30MH074678, P20MH071897, P30MH068685, R01MH06624, R01MH066319, R01MH080122, R01MH078016, R01MH068423, R01MH069353, R34MH071189, R01DA 025192, and the Substance Abuse and Mental Health Services Administration (SAMHSA) under grant number SM57405.

## LITERATURE CITED

1. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc* 1996;91:444–55.
2. Ayanlowo A, Redden D. A two-stage conditional power adaptive design adjusting for treatment by covariate interaction. *Contemp. Clin. Trials* 2008;29:428–38. [PubMed: 18053774]
3. Barnard J, Frangakis CE, Hill JL, Rubin DB. Principal stratification approach to broken randomized experiments: a case study of school choice vouchers in New York City/Comment/Rejoinder. *J. Am. Stat. Assoc* 2003;98:299–323.
4. Bartlett RH, Gazzaniga AB, Jefferies MR, Huxtable RF, Haiduc NJ, Fong SW. Extracorporeal membrane oxygenation (ECMO) cardiopulmonary support in infancy. *Trans. Am. Soc. Artif. Intern. Organs* 1976;22:80–93. [PubMed: 951895]
5. Bartlett RH, Roloff DW, Cornell RG, Andrews AF, Dillon PW, Zwischenberger JB. Extracorporeal circulation in neonatal respiratory failure: a prospective randomized study. *Pediatrics* 1985;76:479–87. [PubMed: 3900904]
6. Begg CB. Investigating therapies of potentially great benefit: ECMO: Comment. *Stat. Sci* 1989;4:320–22.
7. Bekele BN, Shen Y. A Bayesian approach to jointly modeling toxicity and biomarker expression in a phase I/II dose-finding trial. *Biometrics* 2005;61:343–54. [PubMed: 16011680]
8. Bennett CC, Johnson A, Field DJ, Elbourne D, U. K. Collab. ECMO Trial Group. UK collaborative randomised trial of neonatal extracorporeal membrane oxygenation: follow-up to age 4 years. *Lancet* 2001;357:1094–96. [PubMed: 11297963]
9. Berger V, Rezvania A, Makarewicz V. Direct effect on validity of response run-in selection in clinical trials. *Control. Clin. Trials* 2003;24:156–66. [PubMed: 12689737]
10. Berry D. Adaptive trials and Bayesian statistics in drug development. *Biopharm. Rep* 2001;9:1–11.
11. Berry DA. Bayesian statistics and the efficiency and ethics of clinical trials. *Stat. Sci* 2004;19:175–87.
12. Berry, DA.; Mueller, P.; Grieve, A.; Smith, M.; Parke, T., et al. Adaptive Bayesian designs for dose-ranging drug trials.. In: Gatsonis, C.; Carlin, B.; Carriquiry, A., editors. *Case Studies in Bayesian Statistics*. Springer; New York: 2002. p. xivp. 376



13. Bhaumik D, Roy A, Aryal S, Hur K, Duan N, et al. Sample size determination for studies with repeated continuous outcomes. *Psychiatr. Ann.* 2009In press
14. Bloom H. Accounting for no-shows in experimental evaluation designs. *Eval. Rev* 1984;8:225–46.
15. Boekeloo B, Griffin M. Review of clinical trials testing the effectiveness of physician approaches to improving alcohol education and counseling in adolescent outpatients. *Curr. Pediatr. Rev* 2007;3:93–101.
16. Boruch B, Foley E, Grimshaw J. Estimating the effects of interventions in multiple sites and settings: place-based randomized trials. *Socialvetenskaplig Tidskr* 2002;2–3:221–39.
17. Braver SL, Smith MC. Maximizing both external and internal validity in longitudinal true experiments with voluntary treatments: the “combined modified” design. *Eval. Progr. Plann* 1996;19:287–300.
18. Brooner RK, Kidorf MS, King VL, Stoller KB, Neufeld KJ, Kolodner K. Comparing adaptive stepped care and monetary-based voucher interventions for opioid dependence. *Drug Alcohol Depend* 2007;88(Suppl 2):S14–23. [PubMed: 17257782]
19. Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. *BMC Med. Res. Methodol* 2006;8:54. [PubMed: 17092344]
20. Brown CH. Comparison of mediational selected strategies and sequential designs for preventive trials: comments on a proposal by Pillow et al. *Am. J. Community Psychol* 1991;19:837–46. [PubMed: 1793091]
21. Brown, CH.; Faraone, SV. Prevention of schizophrenia and psychotic behavior: definitions and methodologic issues. In: Stone, WS.; Faraone, SV.; Tsuang, MT., editors. *Early Clinical Intervention and Prevention in Schizophrenia*. Humana; Totowa, NJ: 2004. p. 255-84.
22. Brown CH, Indurkha A, Kellam SG. Power calculations for data missing by design with application to a follow-up study of exposure and attention. *J. Am. Stat. Assoc* 2000;95:383–95.
23. 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 Alcohol Depend* 2008;95:S74–104. [PubMed: 18215473]  
Supplementary data, doi:10.1016/j.drugalcdep.2008.01.005
24. Brown CH, Wyman PA, Brinales JM, Gibbons RD. The role of randomized trials in testing interventions for the prevention of youth suicide. *Int. Rev. Psychiatry* 2007;19:617–31. [PubMed: 18092240]
25. Brown CH, Wyman PA, Guo J, Peña J. Dynamic wait-listed designs for randomized trials: new designs for prevention of youth suicide. *Clin. Trials* 2006;3:259–71. [PubMed: 16895043]
26. Bruce M, Ten Have T, Reynolds C, Katz II, Schulberg HC, et al. Reducing suicidal ideation and depressive symptoms in depressed older primary care patients. *JAMA* 2004;291:1081–91. [PubMed: 14996777]
27. Campbell F, Ramey C. Effects of early intervention on intellectual and academic achievement: a follow-up study of children from low-income families program title: Carolina Abecedarian Project. *Child Dev* 1994;65:684–98. [PubMed: 8013248]
28. Chamberlain, P. *Treating Chronic Juvenile Offenders: Advances Made Through the Oregon Multidimensional Treatment Foster Care Model*. Am. Psychol. Assoc.; Washington, DC: 2003. p. xvip. 186ill.; 26 cm
29. Chamberlain P, Brown CH, Saldana L, Reid J, Wang W. Engaging and recruiting counties in an experiment on implementing evidence-based practice in California. *Administration and Policy in Mental Health and Mental Health Services Research* 2008;35:250–60. [PubMed: 18302015]
30. Chappell R, Karrison T, Chappell R, Karrison T. Continuous Bayesian adaptive randomization based on event times with covariates by Cheung et al. *Statistics in Medicine* 2006 2007;25:55–70. *Stat. Med* 26:3050–2. [PubMed: 17163577]author reply pp. 2–4
31. Chow S-C, Lu Q, Tse S-K. Statistical analysis for two-stage seamless design with different study endpoints. *J. Biopharm. Stat* 2007;17:1163–76. [PubMed: 18027223]
32. Cialdini, R. *Influence: Science and Practice*. Allyn & Bacon; Needham Heights, MA: 2001.
33. Collins LM, Murphy SA, Bierman KA. A conceptual framework for adaptive preventive interventions. *Prev. Sci* 2004;5:185–96. [PubMed: 15470938]
34. Collins LM, Murphy S, Nair V, Strecher V. A strategy for optimizing and evaluating behavioral interventions. *Ann. Behav. Med* 2005;30:65–73. [PubMed: 16097907]

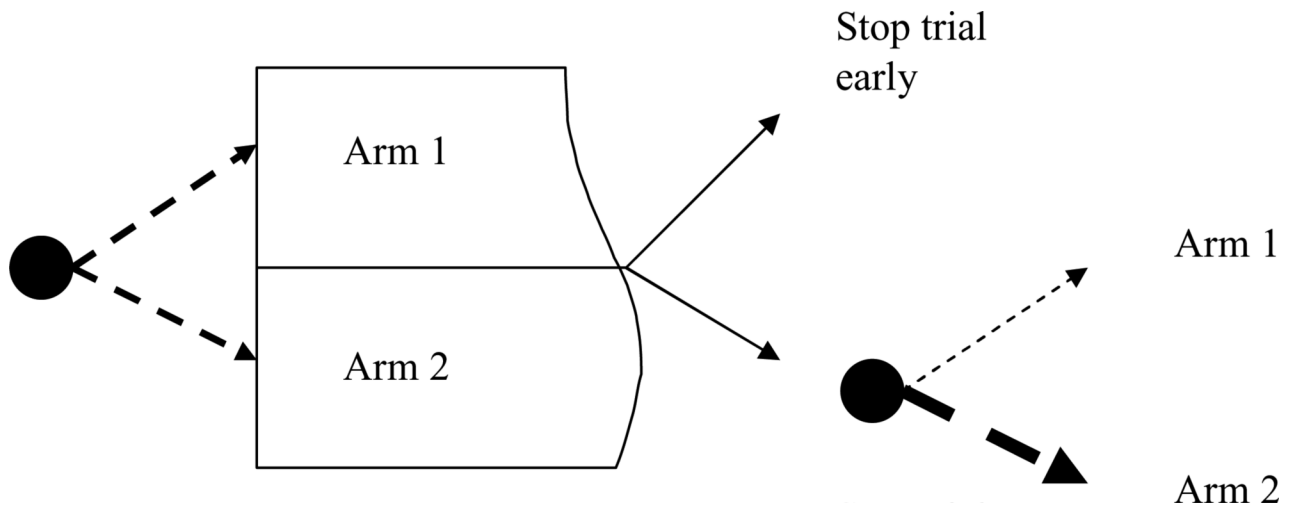
35. Collins LM, Murphy S, Stecher V. The Multiphase Optimization Strategy (MOST) and the Sequential Assignment Randomized Trial (SMART): new methods for more potent eHealth interventions. *Am. J. Prevent. Med* 2007;32:S112–S18.
36. Dexter P, Wolinsky F, Gramelspacher G, Zhou X, Eckert G, et al. Effectiveness of computer-generated reminders for increasing discussions about advance directives and completion of advance directive forms: a randomized, controlled trial. *Ann. Intern. Med* 1998;128:102–10. [PubMed: 9441569]
37. Dolan LJ, Kellam SG, Brown CH, Werthamer-Larsson L, Rebok GW, et al. The short-term impact of two classroom-based preventive interventions on aggressive and shy behaviors and poor achievement. *J. Appl. Dev. Psychol* 1993;14:317–45.
38. Donner, A.; Klar, N. *Design and Analysis of Cluster Randomization Trials in Health Research*. Arnold; London: 2000.
39. Donner A, Klar N. Pitfalls of and controversies in cluster randomized trials. *Am. J. Public Health* 2004;94:416–22. [PubMed: 14998805]
40. Dragalin V. Adaptive designs: terminology and classification. *Drug Inf. J* 2006;40:425–35.
41. Dragalin, V.; Fedorov, VV. Rep. GSK BDS Tech. Rep. 2004–02. GlaxoSmithKlein Pharma.; Collegeville, Penn.: 2004. Adaptive model-based designs for dose-finding studies..
42. Dragalin V, Hsuan F, Padmanabhan SK. Adaptive designs for dose-finding studies based on Sigmoid Emax Model. *J. Biopharm. Stat* 2007;17:1051–70. [PubMed: 18027216]
43. Duan, N.; Wells, K.; Braslow, J.; Weisz, J. Randomized encouragement trial: a paradigm for public health intervention evaluation. Rand Corp.; 2002. Unpublished manuscript
44. Eidelman RS, Hebert PR, Weisman SM, Hennekens CH. An update on aspirin in the primary prevention of cardiovascular disease. *Arch. Int. Med* 2003;163(17):2006–10. [PubMed: 14504112]
45. Ekström E, Hyder S, Chowdhury M. Efficacy and trial effectiveness of weekly and daily iron supplementation among pregnant women in rural Bangladesh: disentangling the issues. *Am. J. Clin. Nutr* 2002;76:1392–400. [PubMed: 12450908]
46. Erickson SJ, Gerstle M, Feldstein SW. Brief interventions and motivational interviewing with children, adolescents, and their parents in pediatric health care settings: a review. *Arch. Pediatr. Adolesc. Med* 2005;159:1173–80. [PubMed: 16330743]
47. Fisher, R. *The Design of Experiments*. Oliver & Boyd; Edinburgh: 1935.
48. Frangakis CE, Brookmeyer RS, Varadhan R, Safaeian M, Vlahov D, Strathdee SA. Methodology for evaluating a partially controlled longitudinal treatment using principal stratification, with application to a needle exchange program. *J. Am. Stat. Assoc* 2004;99:239–50.
49. Frangakis CE, Rubin DB. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* 1999;86:365–79.
50. Frangakis CE, Rubin DB, Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002;58:21–29. [PubMed: 11890317]
51. Frangakis CE, Rubin DB, Zhou XH. Clustered encouragement designs with individual noncompliance: Bayesian inference with randomization, and application to advance directive forms. *Biostatistics* 2002;3:147–64. [PubMed: 12933609]
52. Freedman B. Equipoise and the ethics of clinical research. *N. Engl. J. Med* 1987;317:141–45. [PubMed: 3600702]
53. Freedman DA. Statistical models for causation. What inferential leverage do they provide? *Eval. Rev* 2006;30:691–713. [PubMed: 17093105]
54. Gallo P, Chuang-Stein C, Dragalin V, Gaydos B, Krams M, et al. Adaptive designs in clinical drug development—an executive summary of the PhRMA working group. *J. Biopharm. Stat* 2006;16:275–83. [PubMed: 16724485]discussion 85–91
55. Gallo P, Chuang-Stein C, Dragalin V, Gaydos B, Krams M, Pinheiro J. Adaptive design in clinical research: issues, opportunities, and recommendations—Rejoinder. *J. Biopharm. Stat* 2006;16:311–12.
56. Glasgow R, Vogt T, Boles S. Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *Am. J. Public Health* 1999;89:1323–27.

57. Goetghebeur E, Vansteelandt S. Discussion of ‘Clustered encouragement designs with individual noncompliance: Bayesian inference with randomization, and application to advance directive forms’ by Frangakis, Rubin and Zhou. *Biostatistics* 2002;3:169–71. [PubMed: 12933611]
58. Goldsmith, SK.; Pellmar, TC.; Kleinman, AM.; Bunney, WE., editors. *Reducing Suicide: A National Imperative*. Natl. Acad. Press; Washington, DC: 2002. p. 496
59. Gooley TA, Martin PJ, Fisher LD, Pettinger M. Simulation as a design tool for phase I/II clinical trials: an example from bone marrow transplantation. *Control. Clin. Trials* 1994;15:450–62. [PubMed: 7851107]
60. Gould L. Sample-size re-estimation: recent developments and practical considerations. *Stat. Med* 2001;20:2625–43. [PubMed: 11523073]
61. Gould MS. Suicide and the media. *Ann. N. Y. Acad. Sci* 2001;932:200–21. [PubMed: 11411187] discussion 21–24
62. Gould MS, Greenberg T, Velting DM, Shaffer D. Youth suicide risk and preventive interventions: a review of the past 10 years. *J. Am. Acad. Child Adolesc. Psychiatry* 2003;42:386–405. [PubMed: 12649626]
63. Gould MS, Petrie K, Kleinman MH, Wallenstein S. Clustering of attempted suicide: New Zealand national data. *Int. J. Epidemiol* 1994;23:1185–89. [PubMed: 7721521]
64. Gould MS, Wallenstein S, Kleinman M. Time-space clustering of teenage suicide. *Am. J. Epidemiol* 1990;131:71–78. [PubMed: 2293755]
65. Hahn, J.; Hirano, K.; Karlan, D. Adaptive design of multiple stage experiments using the propensity score. Univ. Munich.; 2007. Unpublished manuscript
66. Hill, A. *Principles of Medical Statistics*. Lancet; London: 1961.
67. Hill, ABS. *Statistical Methods in Clinical and Preventive Medicine*. Livingstone; Edinburgh/London: 1962. p. viii. 610p. 8plates
68. Hirano K, Imbens G, Rubin D, Zhou X. Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* 2000;1:69–88. [PubMed: 12933526]
69. Inoue LYT, Thall PF, Berry DA. Seamlessly expanding a randomized phase II trial to phase III. *Biometrics* 2002;58:823–31. [PubMed: 12495136]
70. Jin H, Rubin DB. Principal stratification for causal inference with extended partial compliance. Application to Efron-Feldman data. *J. Am. Statist. Assoc* 2008;103:101–11.
71. Jo B. Estimation of intervention effects with noncompliance: alternative model specifications. *J. Educ. Behav. Stat* 2002;27:385–409.
72. Jo B. Model misspecification sensitivity analysis in estimating causal effects of interventions with noncompliance. *Stat. Med* 2002;21:3161–81. [PubMed: 12375297]
73. Jo B. Statistical power in randomized intervention studies with noncompliance. *Psychol. Methods* 2002;7:178–93. [PubMed: 12090409]
74. Jo B, Asparouhov T, Muthén BO, Ialongo NS, Brown CH, et al. Cluster randomized trials with treatment noncompliance. *Psychol. Methods* 2008;13:1–18. [PubMed: 18331150]
75. Jo, B.; Muthén, B. Longitudinal studies with intervention and noncompliance: estimation of causal effects in growth mixture modeling.. In: Duan, N.; Reise, S., editors. *Multilevel Modeling: Methodological Advances, Issues, and Applications*. Erlbaum; Mahwah, NJ: 2003. p. 112-39.
76. Jo, B.; Muthén, BO. Modeling of intervention effects with noncompliance: a latent variable approach for randomized trials.. In: Marcoulides, GA.; Schumacker, RE., editors. *New Developments and Techniques in Structural Equation Modeling*. Erlbaum; Hillsdale, NJ: 2001. p. 57-87.
77. Karrison TG, Huo D, Chappell R. A group sequential, response-adaptive design for randomized clinical trials. *Control. Clin. Trials* 2003;24:506–22. [PubMed: 14500050]
78. Kraemer HC, Mintz J, Noda A, Tinklenberg J, Yesavage JA. Caution regarding the use of pilot studies to guide power calculations for study proposals. *Arch. Gen. Psychiatry* 2006;63:484–89. [PubMed: 16651505]
79. Krams M, Burman C-F, Dragalin V, Gaydos B, Grieve AP, et al. Adaptive designs in clinical drug development: opportunities, challenges, and scope reflections following PhRMA’s November 2006 workshop. *J. Biopharm. Stat* 2007;17:957–64. [PubMed: 18027207]

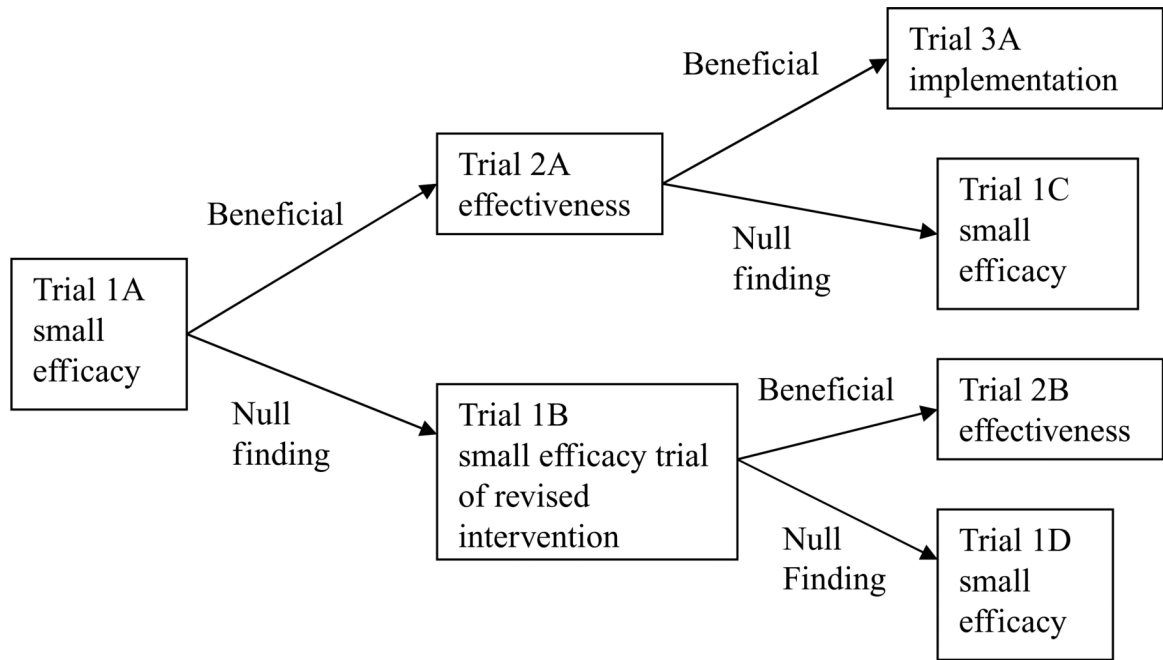
80. Lake S, Kammann E, Klar N, Betensky R. Sample size re-estimation in cluster randomization trials. *Stat. Med* 2002;21:1337–50. [PubMed: 12185888]
81. Lang J. The use of a run-in to enhance compliance. *Stat. Med* 1990;9:87–95. [PubMed: 2345841]
82. Lavori PW, Dawson R. Dynamic treatment regimes: practical design considerations. *Clin. Trials* 1:9–20. [PubMed: 16281458]
83. Lee S, Walker J, Jakul L, Sexton K. Does elimination of placebo responders in a placebo run-in increase the treatment effect in randomized clinical trials? A meta-analytic evaluation. *Depress. Anxiety* 2004;19:10–19. [PubMed: 14978780]
84. Lind, J. *A Treatise of the Scurvy in Three Parts. Containing an Inquiry into the Nature, Causes and Cure of that Disease, Together with a Critical and Chronological View of What Has Been Published on the Subject.* Millar; London: 1753.
85. Liu Q, Pledger G. Phase 2 and 3 combination designs to accelerate drug development. *J. Am. Stat. Assoc* 2005;100:493–502.
86. Lossius MI, Hessen E, Mowinckel P, Stavem K, Erikssen J, et al. Consequences of antiepileptic drug withdrawal: a randomized, double-blind study (Akershus Study). *Epilepsia* 2008;49:455–63. [PubMed: 17888074]
87. McDonald C, Hiu S, Tierney W. Effects of computer reminders for influenza vaccination on morbidity during influenza epidemics. *MD Comput* 1992;9:304–12. [PubMed: 1522792]
88. Mehta C, Tsiatis A. Flexible sample size considerations using information based interim monitoring. *Drug Inf. J* 2001;35:1095–112.
89. Merrell KW, Buchanan R. Intervention selection in school-based practice: using public health models to enhance systems capacity of schools. *Sch. Psychol. Rev* 2006;35:167–80.
90. Mrazek, PJ.; Haggerty, RJ. *Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research.* Natl. Acad. Sci., Inst. Med.; Washington, DC: 1994.
91. Mugford M, Elbourne D, Field D. Extracorporeal membrane oxygenation for severe respiratory failure in newborn infants. *Cochrane Database Syst. Rev* 2008;3:CD001340. [PubMed: 18646070]
92. Muñoz R, Lenert L, Delucci K, Stoddart J, Perez J, et al. Toward evidence-based Internet interventions: a Spanish/English site for international smoking cessation trials. *Nicotine Tob. Res* 2006;8:77–87. [PubMed: 16497602]
93. Murphy SA. An experimental design for the development of adaptive treatment strategies. *Stat. Med* 2005;24:1455–81. [PubMed: 15586395]
94. Murphy SA, Lynch KG, Oslin D, McKay JR, Ten Have T. Developing adaptive treatment strategies in substance abuse research. *Drug Alcohol Depend* 2007;88:S24–30. [PubMed: 17056207]
95. Murray, DM. *Monogr. Epidemiol. Biostat. Oxford Univ. Press; New York: 1998. Design and Analysis of Group-Randomized Trials..*
96. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: race-, sex-, and age-based disparities. *JAMA* 2004;291:2720–76. [PubMed: 15187053]
97. Muthén, BO.; Brown, CH.; Leuchter, A.; Hunter, A. General approaches to analysis of course: applying growth mixture modeling to randomized trials of depression medication.. In: Shrout, PE., editor. *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures.* Am. Psychiatr.; Washington, DC: 2008. In press
98. Muthén BO, Brown CH, Masyn K, Jo B, Khoo ST, et al. General growth mixture modeling for randomized preventive interventions. *Biostatistics* 2002;3:459–75. [PubMed: 12933592]
99. Muthén BO, Jo B, Brown CH. Comment on the Barnard, Frangakis, Hill & Rubin article, “Principal stratification approach to broken randomized experiments: a case study of school choice vouchers in New York City.”. *J. Am. Stat. Assoc* 2003;98:311–14.
100. Nair V, Strecher V, Fagerlin P, Resnicow K, Murphy S, et al. Screening experiments and the use of fractional factorial designs in behavioral intervention research. *Am. J. Public Health* 2008;98:1354–59. [PubMed: 18556602]
101. Natl. Comm. for the Prot. of Hum. Subjects Biomed. and Behav. Res.. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research.* Dep. Health Educ. Welfare; Washington, DC: 1979.

102. O' Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics* 1990;46:33–48. [PubMed: 2350571]
103. O'Rourke PP, Crone RK, Vacanti JP, Ware JH, Lillehei CW, et al. Extracorporeal membrane oxygenation and conventional medical therapy in neonates with persistent pulmonary hypertension of the newborn: a prospective randomized study. *Pediatrics* 1989;84:957–63. [PubMed: 2685740]
104. Pablos-Mendez A, Barr R, Shea S. Run-in periods in randomized trials: implications for the application of results in clinical practice. *JAMA* 1998;279:222–25. [PubMed: 9438743]
105. Perry C, Bishop D, Taylor G, Davis M, Story M, et al. A randomized school trial of environmental strategies to encourage fruit and vegetable consumption among children. *Health Educ. Behav* 2004;31:65–76. [PubMed: 14768658]
106. Pillow DR, Sandler IN, Braver SL, Wolchik SA, Gersten JC, et al. Theory-based screening for prevention: focusing on mediating processes in children of divorce. *Am. J. Community Psychol* 1991;19:809–36. [PubMed: 1793090]
107. Pocock, SJ. *Clinical trials: A Practical Approach*. Wiley; New York: 1983.
108. Proschan MA. Two-stage sample size re-estimation based on a nuisance parameter: a review. *J. Biopharm. Stat* 2005;15:559–74. [PubMed: 16022163]
109. Ridker PM, Cook NR, Lee IM, Gordon D, Gaziano JM, et al. A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. *New Engl. J. Med* 2005;352:1293–304. [PubMed: 15753114]
110. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41–55.
111. Roy A, Bhaumik D, Aryal S, Gibbons R. Sample size determination for hierarchical longitudinal designs with differential attrition rates. *Biometrics* 2007;63:699–707. [PubMed: 17825003]
112. Royall R. Investigating therapies of potentially great benefit: ECMO: Comment. *Stat. Sci* 1989;4:318–19.
113. Royall RM. Ethics and statistics in randomized clinical trials: Rejoinder. *Stat. Sci* 1991;6:83–88.
114. Sandler I, Ostrom A, Bitner M, Ayers T, Wolchik S, Daniels V. Developing effective prevention services for the real world: a prevention service development model. *Am. J. Commun. Psychol* 2005;35:127–42.
115. Sheiner L, Rubin D. Intention to treat analysis and the goals of clinical trials. *Clin. Pharmacol. Ther* 1995;57:6–15. [PubMed: 7828382]
116. Shen Y, Fisher L. Statistical inference for self-designing clinical trials with a one-sided hypothesis. *Biometrics* 1999;55:190–97. [PubMed: 11318154]
117. Smith-Daniels V, Sandler I, Wolchik S. Use of quality management methods in the transition from efficacious prevention programs to effective prevention services. *Am. J. Community Psychol* 2008;41:250–61. [PubMed: 18351452]
118. Sommer A, Zeger S. On estimating efficacy from clinical trials. *Stat. Med* 1991;10:45–52. [PubMed: 2006355]
119. Spiegelhalter, D.; Abrams, K.; Myles, J. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley; New York: 2004.
120. Stuart EA, Perry DF, Le H-N, Ialongo NS. Estimating intervention effects of prevention programs: accounting for noncompliance. *Prev. Sci* 2008;9:288–98. [PubMed: 18843535]
121. Ten Have T, Normand S, Marcus S, Lavori P, Duan N, Brown C. Intent-to-treat vs. non-intent-to-treat analyses under treatment non-adherence in mental health randomized trials. *Psychiatr. Ann* 2008;38(12)In press
122. Ten Have TR, Coyne J, Salzer M, Katz I. Research to improve the quality of care for depression: alternatives to the simple randomized clinical trial. *Gen. Hosp. Psychiatry* 2003;25:115–23. [PubMed: 12676425]
123. Thall PF, Cook JD. Dose-finding based on efficacy-toxicity trade-offs. *Biometrics* 2004;60:684–93. [PubMed: 15339291]
124. Thall PF, Russell KE. A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. *Biometrics* 1998;54:251–64. [PubMed: 9544520]

125. Todd S, Stallard N. A new clinical trial design combining Phase 2 and 3: sequential designs with treatment selection and a change of endpoint. *Drug Inf. J* 2005;39:109–18.
126. UK Collab. ECMO Trial Group. UK collaborative randomised trial of neonatal extracorporeal membrane oxygenation. *Lancet* 1996;348:75–82. [PubMed: 8676720]
127. Ulmer M, Robinaugh D, Friedberg JP, Lipsitz SR, Natarajan S. Usefulness of a run-in period to reduce drop-outs in a randomized controlled trial of a behavioral intervention. *Contemp. Clin. Trials* 2008;29:705–10. [PubMed: 18565801]
128. Victoria CG, Habicht J-P, Bryce J. Evidence-based public health: moving beyond randomized trials. *Am. J. Public Health* 2004;94:400–5. [PubMed: 14998803]
129. Volpp KG, Gurmankin Levy A, Asch DA, Berlin JA, Murphy JJ, et al. A randomized controlled trial of financial incentives for smoking cessation. *Cancer Epidemiol. Biomark. Prev* 2006;15:12–18.
130. Wang L, Cui L. Seamless phase II/III combination study through response adaptive randomization. *J. Biopharm. Stat* 2007;17:1177–87. [PubMed: 18027224]
131. Ware JH. Investigating therapies of potentially great benefit: ECMO. *Stat. Sci* 1989;4:298–306.
132. Ware JH. Investigating therapies of potentially great benefit: ECMO: Rejoinder. *Stat. Sci* 1989;4:337–40.
133. Wei LJ. An application of an urn model to the design of sequential controlled clinical trials. *J. Am. Stat. Assoc* 1978;73:559–63.
134. Wei LJ, Durham S. The randomized play-the-winner rule in medical trials. *J. Am. Stat. Assoc* 1978;73:840–43.
135. Whitehead, J. *The Design and Analysis of Sequential Clinical Trials*. Wiley; New York: 1997.
136. Wolchik, S.; Sandler, I.; Weiss, L.; Winslow, E. New beginnings: an empirically-based intervention program for divorced mothers to help children adjust to divorce.. In: Briesmeister, JM.; Schaefer, CE., editors. *Handbook of Parent Training: Helping Parents Prevent and Solve Problem Behaviors*. Wiley; Hoboken, NJ: 2007. p. 25-66.
137. Wolchik SA, Sandler IN, Millsap RE, Plummer BA, Greene SM, et al. Six-year follow-up of preventive interventions for children of divorce: a randomized controlled trial. *JAMA* 2002;288:1874–81. [PubMed: 12377086]
138. Wood D, Halfon N, Donald-Sherbourne C, Mazel RM, Schuster M, et al. Increasing immunization rates among inner-city, African American children: a randomized trial of case management. *JAMA* 1998;279:29–34. [PubMed: 9424040]
139. Wyman PA, Brown CH, Inman J, Cross W, Schmeelk-Cone K, et al. Randomized trial of a gatekeeper program for suicide prevention: 1-year impact on secondary school staff. *J. Consult. Clin. Psychol* 2008;76:104–15. [PubMed: 18229988]
140. Zelen M. Play the winner rule and the controlled clinical trial. *J. Am. Stat. Assoc* 1969;64:131–46.
141. Zelen M. A new design for randomized clinical trials. *N. Engl. J. Med* 1979;300:1242–45. [PubMed: 431682]
142. Zelen M. Randomized consent designs for clinical trials: an update. *Stat. Med* 1990;9:645–56. [PubMed: 2218168]
143. Zhang X, Cutter G. Bayesian interim analysis in clinical trials. *Contemp. Clin. Trials* 2008;29:751–55. [PubMed: 18589003]
144. Zohar S, Chevret S. Recent developments in adaptive designs for Phase I/II dose-finding studies. *J. Biopharm. Stat* 2007;17:1071–83. [PubMed: 18027217]

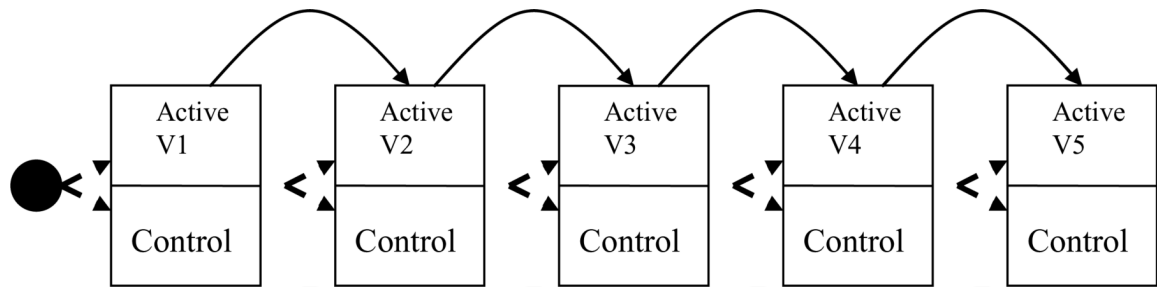


**Figure 1.** Adaptive designs leading to early stopping or modifying allocation probabilities.



**Figure 2.**  
A typical sequencing of trials.





**Figure 3.**  
Cumulative adaptive trials with continuous quality improvement.

Table 1

Adaptive designs	Based on	Example	Benefit	Special trial requirement	Caution	Reference
What is adaptive	Interim analyses of impact	Group sequential trial	Reduced time/cost to completing trial	Requires short time for outcomes to be measured	Wider confidence intervals at the planned end of trial	Whitehead 1997 (135)
Early stopping						
Optimal dose	Reallocation of dosages based on estimated dose-response relationship	Dose finding trials	Efficient determination of dosages	Requires short time for outcomes to be measured	Confounding of dose by entry time	Dragalin & Fedrov 2004 (41)
Sample size	Interim analyses	Reestimation of variance components	Appropriate sample size	Requires short time for outcomes to be measured	Inflated Type I error	Shen & Fisher 1999 (116)
Enrollment procedure	Limited enrollment in traditional randomized design	Randomize before consent	Increased proportion enrolled, decreased selection bias	None	Ethical concerns, especially with single consent design	Zelen 1990 (142)
Run-in prior to randomization	Prediction of drop-out	Run-in design to lower drop-out	Improved statistical power	Multiple responses for baseline	Selection bias, effects of training	Ulmer et al. 2008 (127)
Encouragement	Predictors of adherence	Motivational interviewing	Reduction in participation bias	Incomplete adherence	CACE modeling sensitive to assumptions	Hirano et al. 2000 (68)
Allocation probabilities	Covariates of prior subjects	Covariate adaptive allocation	Enhanced balance in intervention assignment	None	May introduce confounding with time	Wei 1978 (133)
	Responses of prior subjects	Response adaptive design	Reduction in risk to subjects	None	May introduce confounding with time	Karrison et al. 2003 (77)
Follow-up	Response on proximal targets	Two-stage follow-up design	Efficient design for follow-up	Longitudinal follow-up	None	Brown et al. 2007 (24)

Table 2

Adaptive sequences of trials

What is adaptive	Based on	Example	Benefit	Special trial requirement	Caution	Reference
Decision to conduct a full-size trial	Pilot trial	Pilot trial result determines whether intervention is worth testing with a fully powered trial	Avoid funding unworthy programs	Cannot use for long-term outcomes	Sample size and power can be incorrect	Kraemer et al. 2006 (78)
Merging multiple stages of research	Combining research questions	Roll-out trial	Ability to evaluate as a program is implemented	Community wants full implementation and is willing to use wait listing	Challenges in implementing program with sufficient strength	Brown et al. 2006 (25)
Reusing a trial sample in a new trial	Developmental appropriateness of further intervention	Rerandomize subjects to a second trial	Ability to examine early or late intervention	Maintain low attrition	Differential attrition	Campbell et al. 1994 (27)
Screening of intervention components	Fractional factorial designs and analysis	Multiphase optimization strategy	Increased efficiency	Large samples are required	Can fail if there are high-level interactions between components	Brown et al. 2007 (35)
Modification of intervention as it is implemented	Sequence of small experiments that can be combined	Cumulative adaptive trial	Streamlined improvement of the intervention	Sufficiently strong mediators or proximal targets to demonstrate effectiveness	Potential effect on Type I error rate	Brown et al. 2007 (24)