Randomized Controlled Trials

Table of Contents

Authors:
Amy West, Ph.D.
Bonnie Spring, Ph.D.

I. Introduction

Introduction

Behavioral scientist practitioners engage with RCTs in three main ways (Spring, 2007):

- Evidence creators â€" designing and conducting RCTs that generate data
- Evidence synthesizers â€" appraising and integrating RCT evidence, to create a comprehensive understanding of the state of knowledge (i.e., conducting a systematic review)
- Evidence consumers â€" appraising articles produced by evidence creators and

synthesizers in order to apply the knowledge in practice or teaching

In this module, you'll learn:

- what randomized controlled trials (RCTs) are
- the basics of how to design and conduct them

Meet Dr. Austin

Dr. Austin is a clinical psychologist who specializes in treating adolescents with depression. She works in a psychiatric clinic in an academic medical setting. Dr. Austin has developed a new treatment for her patients. Her treatment combines a few techniques that have previously been shown to be effective with ideas of her own about what might be helpful.

Dr. Austin started delivering her new treatment, and soon she observed that her patients seemed to get better. To learn whether her impressions were accurate, Dr. Austin got approval from her university and consent from her adolescent patients and their parents to measure the patients' symptoms and functioning before and after they received the treatment.

Dr. Austin's measurements showed that, after receiving her treatment, the adolescents reported that their symptoms had improved and they were functioning better in their lives.

Dr. Austin was pleased, but she couldn't be sure whether the improvement was actually caused by her new treatment or by other influencesâ€”such as just getting attention from a therapist or the natural remission of symptoms over time. When she discussed this problem with a colleague, the colleague suggested that Dr. Austin conduct an RCTâ€”a randomized controlled trial.

What is a Randomized Controlled Trial (RCT)?

Randomized Controlled Trial

- A controlled trial is a study in which participants are assigned to a study group. Study groups are also called study arms or treatment conditions.
- In a randomized controlled trial, participants are assigned to treatment conditions at random (i.e., they have an equal probability of being assigned to any group).
- Procedures are controlled to ensure that all participants in all study groups are treated the same except for the factor that is unique to their group. The unique factor is the type of intervention they receive.

The primary goal of conducting an RCT is to test whether an intervention works by comparing it to a control condition, usually either no intervention or an alternative intervention. Secondary goals may include:

- identify factors that influence the effects of the intervention (i.e., moderators)
- understand the processes through which an intervention influences change (i.e., mediators or change mechanisms that bring about the intervention effect)

Brief History of RCTs

- Descriptions of experiments date back to biblical times. However, the first instance of random allocation of patients to experimental and control conditions is attributed to James Lind, a naval surgeon, in 1747.
    - Lind randomly assigned 12 sailors to 6 different candidate treatments for scurvy. The two patients who were given lemons and oranges recovered most quickly, suggesting a beneficial effect of citrus. Question to consider...

- The first RCT in medicine is credited to Sir A. Bradford Hill, an epidemiologist for England's Medical Research Council. The trial, published in the British Medical Journal in 1948, tested whether streptomycin is effective in treating

tuberculosis. Because the drug was in short supply, Hill devised randomization partly as a strategy to keep doctors from trying to maneuver their patients into the streptomycin arm of the trial. Earlier in his career, Hill believed that simply alternating the assignment of hospital admissions to drug versus control worked well enough. Later he recognized that simple alternation led to selection bias because the sequence was too easy to predict. That realization led to the use of a random numbers table to generate the numeric series by which patients would be assigned to conditions.

- Over the years, use of RCTs has increased in the medical, social, and educational sciences. The change reflects a growing recognition that observational studies without a randomly assigned control group are a poor way of testing whether an intervention works.

Question to consider

Do you think that 2 participants per treatment group would be considered sufficient in a contemporary RCT?

Why Conduct an RCT?

An RCT is conducted to test whether an intervention or treatment works. The key methodological components of an RCT are (1) use of a control condition to which the experimental intervention is compared; and (2) random assignment of participants to conditions. Advantages of using an RCT design include:

- Random assignment ensures that known and unknown person and environment characteristics that could affect the outcome of interest are evenly distributed across conditions.
- Random assignment equalizes the influence of nonspecific processes not integral to the intervention whose impact is being tested. Nonspecific processes might include effects of participating in a study, being assessed, receiving attention,

self-monitoring, positive expectations, etc.
- Random assignment and the use of a control condition ensure that any extraneous variation not due to the intervention is either controlled experimentally or randomized. That allows the study's results to be causally attributed to differences between the intervention and control conditions.

In sum, the use of an RCT design gives the investigator confidence that differences in outcome between treatment and control were actually caused by the treatment, since random assignment (theoretically) equalizes the groups on all other variables.

Misconceptions and Drawbacks to RCT Methods

Common misconceptions about RCTs:

- RCTs are sometimes criticized mistakenly for not being practical or generalizable. This criticism arises because initial studies to develop a treatment or test its efficacy are often conducted in a controlled setting (e.g., academic research center), with well-trained interventionists, and patients who have only the single problem being studied. However, once a treatment has shown efficacy, later effectiveness RCTs evaluate generalizability.
- Some people have ethical concerns about denying the control group an intervention that they *believe* will be helpful. However, the reason for doing the RCT is that it isn't truly known whether the new treatment will help, harm, or have no effect.

True drawbacks of conducting an RCT are:

- They are time- and energy- intensive
- They are expensive
- They may not be feasible for all interventions or settings (e.g., some institutions have policies that prohibit random assignment)

How Are RCTs Used in Behavioral Sciences?

The most common use for RCTs in the behavioral and social sciences is to examine whether an intervention is effective in producing desired behavior change, symptom reduction, or improvement in quality of life.

Having consistent findings that the intervention surpasses control in a series of RCTs is often considered to establish the intervention as "evidence-based" (i.e., that it has sufficient data to support its use).

For example...

For example...

Cognitive-behavioral therapy (CBT) is a style of therapy that focuses on changing troublesome thoughts, feelings, and behavior. CBT for the treatment of an anxiety disorder has been studied extensively via RCTs.

A systematic review combining evidence from many RCTs conducted in different settings, with different populations, and somewhat different protocols was commissioned by the Cochrane Collaboration. The review found remission of an anxiety disorder in 56% of anxious children treated with CBT - twice the remission rate found in controls (James, Soler, Weatherall, 2005). As a result of the accumulation of high-quality evidence in this area, CBT is now considered the front-line treatment for childhood anxiety disorders.

II. Designing an RCT

Maximizing Validity and Minimizing Bias

The usefulness of a trial depends on the extent to which it lets us validly infer that the experimental treatment caused an outcome. The ability to make valid inferences depends on how well the investigator designed, conducted, and reported various procedures to minimize bias in the study.

- **Bias**

  Bias is a systematic distortion of the real, true effect that results from the way a study was conducted. This can lead to invalid conclusions about whether an intervention works. Bias in research can make a treatment look better or worse than it really is. Inferences about validity fall into four primary categories: internal, external, statistical conclusion, and construct validity.

- **Internal Validity**

  Internal validity is the extent to which the results of a study are true. That is, the intervention really did cause the change in behavior. The change was not the result of some other extraneous factor, such as differences in assessment procedures between intervention and control participants.

- **External Validity**

  External validity is the extent to which the results can be generalized to a population of interest. The population of interest is usually defined as the people the intervention is intended to help.

- **Statistical Conclusion Validity**

  The validity of inferences about covariation between two variables.

- **Construct Validity**

  The extent to which the study tests underlying constructs as intended.

Maximizing Validity and

Minimizing Bias

The question here is: to what extent can the intervention, rather than other influences, be considered to account for differing outcomes between the treatment and control groups?

By employing a control group and random assignment, the RCT design minimizes bias and threats to internal validity by equalizing the conditions on all "other influences" except for the intervention. Here are common threats to internal validity that random assignment addresses.

- **History:** external events that occur during the course of a study that could explain why people changed. An example is a life event such as a death in the family or being laid off. Negative life events, such as these, may offer an alternative explanation for study outcomes. Consequently, it is very useful that random assignment equalizes these occurrences across the treatment and control conditions.

  For example...

- **Maturation:** processes that occur within individuals over the course of study participation that provide an alternate explanation of why they changed. For example, depression increases after the onset of puberty. Thus, if more children in the control than the treatment group reached puberty during the course of the study, that might explain why the control group finished the study with more depression than the treated group.
- **Temporal Precedence:** in order to establish a causal relation between the intervention and outcome, the intervention must occur before the outcome.
- **Selection:** another threat to validity may occur if the experimental and control groups differ at baseline.
- **Regression to the Mean:** this threat refers to the tendency for extreme scorers to regress to the mean on subsequent measurements.
- **Attrition:** refers to a rate of loss of participants from the study that differs between the intervention and control groups.
- **Testing and Instrumentation:** these threats refer to changes due to the nature of measurement rather than the intervention itself.

For example...

A factory shutdown occurred during the course of Dr. Austin's study, causing many town residents to lose their jobs. Consider a scenario in which all of the children in Dr. Austin's control group and none of the children in her treatment group had a parent laid off in the shutdown. If the treated children were less depressed than the control children at the end of the study, Dr. Austin would not know what caused the difference. It might have been her treatment. Just as plausibly it might have been that the children in her treatment experienced a lower number of stressful life events than the control children. Fortunately, randomizing patients to conditions increases the probability that the treatment and control groups will have similar exposure to extraneous events.

Minimizing Threats to External Validity

The question here is: to what extent can the results be extended to people, settings, and interventionists different than those used in this particular study?

These are common threats to external validity that an RCT can address.

- **Sample Characteristics:** the extent to which we can generalize from the study sample to the population as a whole (or the population of interest). External validity can be enhanced by having the study sample include representation of a range of important demographic characteristics (e.g., gender, ethnicity, SES). A broadly representative sample enables the findings to be generalized to a diverse population. It may also allow the investigators to explore whether the treatment appears more or less effective with some population subgroups.
- **Setting Characteristics:** the ability to generalize beyond the particular setting in which the study is conducted (e.g., clinic, therapists, study personnel). Concern about external validity can arise when a study intervention is delivered by highly trained research staff in an academic setting. The question is whether the intervention could be applied by less professionally credentialed staff in an under-resourced setting.

  Investigators must balance the need for control, rigor, and efficiency with a desire to have results be relevant to the broad range of settings and populations they could potentially benefit. This choice often hinges on the stage of intervention development and where the particular study falls on the

spectrum from testing efficacy to effectiveness.

- **Effects Due to Testing:** refers to the potential for participants to respond differently because they know they are being assessed as part of research.

Testing Efficacy vs. Effectivenessa>

An efficacy trial answers the question: "Does this intervention work under optimal conditions?" An effectiveness trial answers the question: "Does this intervention work under usual conditions?"

Efficacy trials are sometimes called explanatory trials, whereas effectiveness trials are also known as pragmatic trials.

For example...

For example...

Dr. Austin mulls over whether to conduct a pragmatic or an explanatory trial. She would like her findings to generalize to the broadest possible population of depressed adolescents, so she leans toward doing a pragmatic trial. However, she has doubts about whether the treatment will work with adolescents who have substance abuse disorders or symptoms of antisocial personality. After talking with a statistician, she realizes that she would need a much larger sample size to conduct a pragmatic trial, because she expects the treatment effect to be less consistent than it would be in an efficacy trial. She decides that her first priority is to learn whether her treatment can work, and conducts an efficacy trial.

Pragmatic vs. Explanatory

Very few trials actually fall wholly into one of these categories, but rather fall along a continuum of pragmatic-explanatory.

Where the study falls on the continuum will depend on many factors, including the stage of intervention development, research question, and resources available.

The stage of intervention development plays an important role. In the beginning stages, an investigator usually wants to know whether the treatment can work and, therefore, is worth developing further. The investigator may opt to keep sample size manageable by conducting a very tightly controlled efficacy study.

However, later in development, once the treatment's efficacy has been established, an investigator may want to know whether the intervention can be more broadly applied in other settings and delivered by nonprofessionals who have less experience with the treatment.

Pragmatic vs. Explanatory

The PRECIS (Thorpe et al, 2009) is a tool that can help investigators design an RCT that falls where they want it to on the continuum between pragmatic and explanatory.

PRECIS itemizes key parameters about which investigators will make different design decisions depending on whether they are planning an explanatory or a pragmatic trial. Entry and exclusion criteria, the expertise and training of interventionists, and the degree of flexibility they are allowed in administering the treatment differ in the two kinds of trials.

Critical Aspects of the Randomized Controlled Trial Design

You know that randomized controlled trials provide the most effective way to control for extraneous influences when testing whether a treatment works. In this module, you'll learn more about the important considerations that are involved in designing an RCT.

We will explore the following important aspects of RCT design:

- Sample selection
- Choice of a control condition
- Random assignment
- Blinding
- Planning for assessment and data collection
- Intervention/Treatment integrity

Sample Selection - Introduction

In this section, you will learn more about sample selection in randomized controlled trials, including:

- Important questions to consider in sample selection
- Selection bias
- Inclusion/exclusion criteria

Sample Selection

The sample selected for the study should as closely approximate the population of interest as possible. When designing the study, it is important for investigators to consider such questions as:

- How will the target population be identified (e.g., What will selection criteria be? Where will recruitment occur and who will be involved)?
- Who will be invited to participate?
- What will the inclusion and exclusion criteria be?
- How will eligibility be assessed?
- Will the selected sample be representative of the population to which I want to generalize?
- How many eligible participants do I expect to need to screen in order to randomize the sample size I need?

For example...

For example...

An investigator wants to test the effectiveness of a school-based violence prevention intervention. He identifies his population of interest as public school children in grades 9-12 in the city where he lives, whose parents allow them to participate.

- **How will the target population be identified (e.g. where will recruitment occur and who will be involved)?**
  - The investigator needs to think about how potential participants will be identified. Will teachers identify students? Will parents refer their children? Should the investigator send study staff to the school to do the recruitment or should he rely on school personnel?
- **Who will be invited to participate? What will be the inclusion and exclusion criteria?**
  - Will all children at the target schools be invited to participate? Or will enrollment be restricted to only students with risk factors for violence or a history of violent behavior? Might selective enrollment stigmatize participants? Will any students be excluded, including the most violent teens?
- **How will eligibility be assessed?**

- Who will decide which students are eligible? How will this be assessed and documented? Will teachers be asked to screen students for eligibility? Will they have time to do that?

- **Will the sample selected be representative of the population the investigator wants to reach?**
  - If the investigator invites children with a certain cluster of risk factors for violence to participate, will the resulting sample reflect the broader population he wants to serve? How will he ensure an adequate number of girls and ethnic minorities (if he wants the intervention to help these groups)?

Subject Selection Bias

Selection bias is a systematic distortion of evidence that arises because people with certain important characteristics were disproportionately more likely to wind up in one condition. Although random assignment theoretically eliminates selection biases, a bias can still occur.

For example...

For example...

Dr. Jones conducted a placebo-controlled RCT testing the efficacy of a new medication to prevent recurrent heart attacks. Although he randomized patients to drug versus placebo, more overweight people were assigned to his placebo condition. Because being overweight heightens the risk of heart attack, a lower heart attack rate among drug- than placebo-treated patients could occur for one of two reasons. The medication that the patients received could have lowered the risk of heart attack. Alternatively, patients randomized to the drug treatment group could have been less at risk because they were less overweight.

Subject Selection Bias

Selection bias threatens the equivalency of the groups and means that the randomization was unsuccessful at balancing important variables across conditions. Important variables are any that relate theoretically or statistically to the study's outcome. If selection bias occurs, it is usually due to chance rather than intentional bias. Nevertheless, investigators need to measure these characteristics at baseline to evaluate their equivalency across conditions.

The risk of selection bias can be reduced, though not eliminated, by stratifying study candidates on relevant characteristics (e.g., gender, age, setting) and then randomizing them to conditions by strata. Stratified randomization is discussed in more detail in the section on random assignment.

Inclusion/Exclusion Criteria

Inclusion and exclusion criteria are criteria that an investigator develops before beginning the study that will define who can be included and who will be excluded from the study sample.

Inclusion and exclusion criteria should:

- Directly relate to the research question being answered
- Be restrictive enough to narrow the sample to one that is clearly defined while being broad enough to generalize to the population of interest

It is important to choose reliable and valid measures of inclusion and exclusion criteria so that the study sample validly reflects the population of interest.

Control Condition - Introduction

In this section, you will learn more about the control conditions in randomized controlled trials, including:

- Different kinds of control conditions
- Considerations when choosing a comparison condition

NOTE: the terms "control condition" and "comparison condition" are used interchangeably.

Control Group Comparisons

There are many alternative control conditions. None is perfect or suitable for all occasions. The choice of a control condition usually depends on the specific question being asked and the state of existing knowledge about the intervention under study.

- **No-treatment Comparison Condition**

  In this comparison, outcomes for people randomly assigned to receive the new treatment are compared to those of people assigned to receive no treatment at all. The question is whether the new treatment produces any benefit at all, over and above change due to the passage of time or the effects of participating in a study. A challenge with this control condition is that people randomized to no treatment may find their own treatment outside the bounds of the study.

- **Wait-list Comparison**

  In this comparison, people randomized to receive the new treatment are compared to those randomized to be on a wait-list to receive the new treatment. Using a wait-list control has the advantage of letting everyone in the study receive the new treatment (sooner or later). A limitation is that expectations of improvement differ between the treatment and control group. The control group knows that they are not yet receiving an active treatment and has no reason to expect positive change. Other possible threats are that people content to sit on a waiting list may be atypical (unusually cooperative), or they may seek other "off-study" treatments on their own.

- **Treatment as Usual Comparison (TAU)**

  In this comparison, people randomized to receive a new treatment are compared to those randomized to receive treatment as usual (i.e., whatever intervention is standard practice). Treatment as usual helps to equalize groups on the expectation of benefit since both groups receive an intervention, although those randomized to the new intervention may still expect something special. However, treatment as usual is particularly well-suited to answer the practical question of whether introducing the new treatment could improve outcomes over and above the current state of practice.

- **Attention/Placebo Comparison**

  In this comparison, the new treatment is compared to a control intervention that delivers the same amount of support and attention from a practitioner, but none of the key active intervention ingredients by which the new treatment is expected to cause change in the outcomes under study. Given evidence that a treatment works, compared to an â€œeasierâ€□ control condition (e.g., no treatment, wait-list, TAU), the attention control tests whether the new treatment produces benefits beyond the effects due to nonspecific influences â€" like therapist attention or positive expectations.

- **Relative Efficacy/Comparative Effectiveness**

  This approach involves a "head-to-head" comparison between two or more treatments, each of which is a contender to be the best practice or standard of care. To detect a difference between conditions, comparative effectiveness trials require many, many participants in each treatment group. That is because all of the interventions being compared are known to work, so the expected difference between them is relatively small. The questions in comparative effectiveness are usually: (1) which intervention works better?; and (2) at what relative costs? Some countries and some insurers use comparative effectiveness findings to determine which treatments to pay for.

- **Parametric/Dose Finding**

  Parametric studies are usually done early in the development of a new treatment in order to determine the optimal "dose" or format of treatment. Different forms of the intervention varying on factors such as the number, length, or duration of sessions comprise the conditions to which people are randomly assigned.

- **Additive/Constructive Comparison**

  Similar to a parametric study, in an additive/constructive comparison, different randomized groups receive different versions of the treatment. Those in the experimental condition receive added treatment components that are

hypothesized to add efficacy. An additive trial may be conducted early in treatment development or after a treatment is well-established, to see if its efficacy can be improved even further.

- **Treatment Dismantling**

  Also called "component analysis," in this approach, people randomized to receive the full efficacious intervention are compared to those randomized to receive a variant of that intervention minus one or more parts. Dismantling designs are usually used late in a treatment's development, after the intervention's efficacy is well-established. The purpose is to determine which components are essential and which may be superfluous. One variant of a dismantling design aims to find the "MINC" â€“ the minimum intervention needed to produce change. The aim, from a public health perspective, is often to find a low-cost, minimally intensive intervention that improves outcomes for a small percent of the population, which equates in absolute numbers to a large number of people being helped.

Random Assignment - Introduction

In this section, you will learn more about random assignment in RCTs, including:

- Goals of randomization
- Randomized versus non-randomized assignment
- Different types of RCT designs
- Fixed and adaptive random assignment procedures
- Allocation concealment

Random Assignment

The goal of randomization is to produce study groups that are comparable on known

and unknown extraneous influences that could affect the study outcome. Randomization achieves this goal by giving all participants an equal chance of being in any condition.

In general, the preferred method of randomization involves a 3rd party (i.e., someone not involved in any other way with the study) generating numbers from a table or computer program. Such a process eliminates even unconscious bias from the random assignment process.

Randomization should occur as close as possible to the initiation of the intervention. This prevents randomizing participants who drop out before participating in any of the study. This is important because everyone who gets randomized needs to be included in the study's analyses.

Randomized Versus Non-randomized Designs

Sometimes randomization is not possible, and it is necessary to consider alternative designs. In quasi-experimental designs, participants are assigned to a study condition using some non-random (but systematic) procedure. Some examples of quasi-experimental designs are:

- Time-series design
- Counterbalanced design

We'll discuss these designs later. Bear in mind, though, that non-random assignment heightens the risk of bias, so random assignment is preferable.

Different Types of RCT Design

There are several variations on the basic RCT design.

- **Cross-over Design**

  This describes a special case of a randomized controlled trial wherein each subject serves as his/her own control.

  - In this design, a study is divided into two time periods:
    - During the first time period, each participant receives either the control treatment or the experimental treatment
    - During the second time period, participants switch conditions
  - The initial treatment each subject receives is determined by random assignment

- **Group Randomized Design**

  The group or cluster randomized design describes an approach whereby whole groups of participants (e.g., schools, clinics, worksites) are randomized to intervention or control. The unit of randomization is a group rather than an individual. This design is often used in situations where there is concern about contamination.

  Contamination occurs when individuals randomized to the intervention condition and those randomized to control are exposed to the wrong condition through having contact with each other. Contamination can occur either inadvertently or intentionally as people discuss their experiences. The cost to internal validity is that people in the control condition receive part of the intervention.

  Group randomization reduces the likelihood of contamination. However, it introduces the problem that settings can have unique properties whose influences become confounded with treatment assignment.

  For example...

For example...

A researcher wants to test a teacher-focused intervention in schools. However, she worries that if she randomizes teachers within a school to receive the intervention or

not, there may be contamination. She fears that teachers will talk to each other about the study, and that control group teachers will observe intervention teachers' behavior and model it, even though they were not randomized to receive the intervention. In response, the researcher decides to randomize different schools to receive the teacher intervention or no intervention.

Fixed Randomization Procedures

In fixed allocation randomization, each participant has an equal probability of being assigned to either treatment or control and the probability remains constant over the course of the study. That can be achieved by using a table of random digits or randomization software (in SAS, SPSS, and other major software programs).

- **Simple (Complete) Randomization**

  This refers to the most elementary form of randomization, in which, every time there is an eligible participant, the investigator flips a coin to determine whether the participant goes into the intervention or control group.

  A limitation is that random assignment is truly random. A random process can result in the study winding up with different numbers of subjects in each group. This is more likely to happen if sample size is small.

- **Blocked Randomization and the Method of Randomly Permuted Blocks**

  Blocked randomization reduces the risk that different numbers of people will be assigned to the treatment (T) and control (C) groups. Patients are randomized by blocks. For example, with a fixed block size of 4, then patients can be allocated in any of the orders: TTCC, TCTC, CTCT, TCCT, CTTC, or CCTT. The order is chosen randomly at the beginning of the block. In randomly permuted blocks, there are several block sizes (e.g., 4, 6, and 8), and the block size and specific order are chosen randomly at the beginning of each block.

  - Blocked randomization offers the advantage that at any point in the trial, there will be a balance in the number of cases assigned to T versus C (which could be valuable if the trial needs to be stopped early).
  - A disadvantage is that, with fixed blocks, research staff may be able to predict the group assignment of patients being randomized late in the

block. That risk is reduced by using the method of randomly permuted blocks and keeping research staff blind to the randomization process.

- **Stratified Randomization**

  It may be important to ensure that the treatment and control groups are balanced on important prognostic factors that can influence the study outcome (e.g., gender, ethnicity, age, socioeconomic status). Before doing the trial, the investigator decides which strata are important and how many stratification variables can be considered given the proposed sample size. A separate simple or blocked randomization schedule is developed for each stratum.

  Large trials often use randomly permuted blocks within stratification groups. This assures that treatment assignments are balanced at the end of every strata block. However, this approach is complex to implement and may be inappropriate for smaller trials.

Fixed Randomization Procedures

In fixed allocation procedures, the probability of being assigned to any treatment stays constant over the course of the trial. In adaptive procedures, the allocation probability changes in response to the balance, composition, or outcomes of the groups. Adaptive randomization procedures remain controversial because they allocate patients not purely at random but partly dependent upon what has already occurred in the trial. The aim of adaptive procedures is efficiently to increase the sample's probability of being assigned to the best treatment.

There are two forms of adaptive procedures.

- **Minimization Adaptive Randomization**

  Minimization corrects (minimizes) imbalances that arise over the course of the study in the numbers of people allocated to the treatment and control.

  - **Biased Coin Randomization**
    - In this procedure, if the imbalance in treatment assignments passes some threshold, the allocation is changed from chance to a bias in favor of the under-represented group. For example...
  - **Urn Randomization**
    - This procedure tries to correct imbalances after each allocation. For example...

- **Response Adaptive Randomization**

  In minimization, a knowledge of the number (and sometimes the strata) of those already randomized shapes decisions about how the next participants will be allocated. In responsive adaptive randomization, knowledge of how the allocated participants have responded to the interventions influences the next allocation probability.

  - **"Play-the-winner" procedure:**
    - In this procedure, like urn randomization, the first allocation is chosen at random from two different colored balls. If the participant has a positive response to the treatment, a ball of the same color is added to the urn. If the treatment fails, a ball of the opposite color is added. This procedure is meant to maximize the number of participants receiving the "superior" intervention. Adaptive methods were developed in response to ethical concerns about assigning participants to conditions that are known to be inferior. A challenge is that the outcome of each treatment is often not known at the time that the next patients need to be assigned.

For example...

If after 10 randomizations, there are 7 patients assigned to intervention and 3 assigned to control, the coin toss will become biased. Then, rather than having 50/50 chance of being assigned to either condition, the next patient will be given a 2/3 chance of being assigned to the under-represented condition and a 1/3 chance of being assigned to the overrepresented one. This procedure requires keeping track of imbalances throughout the trial. In smaller trials, imbalances can still result.

For example...

The investigator starts with off with an urn containing a red ball and a blue ball to

represent each condition. If the first draw pulls the red ball, then the red ball is replaced together with a blue ball, increasing the odds that blue will be chosen on the next draw. This continues, replacing the chosen ball and one of the opposite color on each draw. The procedure works best at preventing imbalance when final sample size will be small.

Allocation Concealment

Preferably, randomization should be completed by someone who has no other study responsibilities, because otherwise their knowledge of the patient's assignment could introduce bias. Often, the study statistician assumes responsibility for performing the randomization. In multi-site trials, randomization usually occurs at a centralized location.

Allocation concealment means that the person who generates the random assignment remains blind to what condition the person will enter. If allocation is not concealed, research staff is prone to assign "better" patients to intervention rather than control, which can bias the treatment effect upward by 20-30% (Wood, 2008).

Blinding - Introduction

In this section, you will learn more about blinding in randomized controlled trials, including:

- Purpose of blinding
- Levels of blinding

Purpose of Blinding

In blinding, the researchers collecting data are prevented from knowing certain information about a participant (e.g., what condition they are in) in order to prevent this information from affecting how they collect data.

Ideally, to minimize bias, both the participant and the investigator are kept blind to (ignorant of) the participant's random assignment. That level of blinding (or masking) may or may not be feasible.

Investigators should implement the greatest level of blinding that is feasible.

Levels of Blinding

In the case of clinical trials, there are several levels of blindness to consider.

- **Double-blind Trials**

  In double-blinding, neither the participants nor the investigator know the participants' treatment assignment. This level of blinding reduces the influence of expectations held by participants or by research staff about which treatment will have a better effect on the outcome.

  - Double-blinding is most feasible for drug trials, in which the effect of a medication is being compared to a placebo that looks similar.
  - Even in placebo-controlled trials, guesswork about the drug assignment is often better than chance. Masking can be improved by using an active placebo that has the same side effects as the drug but lacks its therapeutic effects.
- **Partial Blinding**

  Double-blinding is rarely possible in trials of behavioral treatment. It is usually obvious to participants which treatment they are receiving. Also, the treatment assignment is known by any research staff who delivers the treatment. However, the staff who assess the study outcome can and should be kept blind to the patient's treatment condition.

  - Special care is needed to prevent staff and study participants from

unblinding the outcome assessor.

- **Unblinded Trials**

  Especially when neither participant nor investigator can be blinded, it is best if participants and research staff hold equally positive expectations about the merits of the treatment and control conditions.

  - The state of equipoise, uncertainty about which intervention condition will work best, is also the ethical justification for conducting a trial.

Assessment and Data Collection - Introduction

In this section, you will learn more about assessment and data collection in randomized controlled trials, including:

- Outcome measurement
- Reliability and validity of outcome measures
- Quality control in data collection

Outcome Measurement

No matter how well-designed, an RCT is only as good as its outcome assessment. It is critically important that investigators think through and specify in advance the outcomes they plan to measure to test whether their treatment works. Ideally, there should be only 1 primary outcome and perhaps 1-2 secondary outcomes. Those outcomes need to be measured as accurately as possible.

- The outcome may be an event (or endpoint), such as death or hospitalization, or the onset or remission of a condition, like major depression.

Outcome Measurement

The term intermediate endpoint or surrogate marker is sometimes used to designate an outcome that is correlated with but not identical to a clinical endpoint.

- Biomarkers (like blood pressure, lipids, or obesity) or health risk behaviors (like smoking, eating a high fat diet, or being physically inactive) can be considered intermediate markers because they relate to disease.

Outcome Measurement

Bias is less when an outcome can be measured objectively.

For example...

Resource utilization and staff time can be monitored to measure the cost of implementing a new treatment.

Many of the outcomes we measure in behavioral clinical trials are subjective (known only to the individual) and need to be measured by self-report. Symptoms of anxiety and perceived quality of life illustrate two outcomes that need to be self-reported.

For example...

Neuroimaging and biomarker data can objectively track the course of a health condition. A device called a MEMS cap, which records the opening of a prescription bottle, measures medication compliance objectively. An accelerometer that counts movements directly measures physical activity.

Reliability of Outcome Measures

Reliability refers to the consistency or repeatability of a measure. It is usually measured in three ways:

- temporal stability
- internal consistency
- equivalence

Validity of Outcome Measures

Validity describes how well a test measures what it is intended to measure.

- Validity is assessed by:
    - how well the test correlates with other measures of the construct
    - how well the test shows predicted relationships to measures of other theoretically related constructs

For example...

For example...

Q: If a person scores as depressed on a questionnaire but doesn't qualify for a depression diagnosis on a semi-structured interview, does the condition of depression exist in that individual?

A: It depends on which is the more valid measure. A semi-structured interview usually constitutes the gold standard for diagnosing a psychological condition. Diagnostic interviews are used to validate questionnaire measures. So, in this case, in the absence of other valid evidence suggesting depression, we would conclude that the person is not currently depressed.

Sensitivity and Specificity of Outcome Measures

When a measure classifies people into one category or another (e.g., sick or not sick), its quality can be evaluated by its:

- sensitivity
- specificity

Data Collection and Quality Control

All data must be:

- high quality
- directly pertinent to the main aims of the study
- collected in a systematic and closely monitored manner

- **Problems in Data Collection**

  Problems that occur in data collection include:

  - erroneous data
  - missing or incomplete data
- **Minimizing Poor Quality Data**

  Actions that investigators can take to ensure the quality of their data include:

  - design a clear, comprehensive study protocol and manual of operations that guides research staff in implementing the study
  - create clear and concise study forms
  - pre-test the protocol and forms
  - train and certify staff in order to standardize implementation of the protocol
  - perform periodic blind assessments of study staffs' performance of study procedures, with retraining as needed to minimize error and drift
  - careful data entry
- **Quality Monitoring**

  Ongoing quality monitoring is necessary to detect errors and missing data in a timely manner that allows them to be corrected. Quality monitoring is helped by having:

  - a study data manager who reviews data from each assessment, contacts study staff to correct erroneous or missing values, and follows up to ensure timely compliance
  - regular audits with feedback from the data manager to study staff about protocol adherence and missing/erroneous data
  - periodic blind reviews of data ranges by the data manager to detect anomalies that might suggest flawed assessment tools
  - study team members code each others' protocol adherence and treatment fidelity
- **Participant Retention**
  - Some missing data are inevitable in a clinical trial; the goal is to minimize them as much as possible. Certain steps can be taken to encourage full participation in outcome assessments:

    - Shorter, simpler study protocols that minimize participant burden tend to reduce drop-out. Burden increases as participants need to

attend frequent in-person visits, take many tests, or undergo unpleasant study procedures (e.g., rectal probes).

- A thorough informed consent process that makes clear what is expected encourages participants to evaluate realistically whether they should enroll.
- A positive relationship with study staff is critically important for retaining participants.
- Support for study involvement from peers and family helps retain participants.
- Because missing data truly convey no information and require making unverifiable assumptions, it is very important for research staff to elicit as much follow-up data as possible, even from participants who decline further participation in treatment.
- Ongoing monitoring of missing data helps investigators understand and correct reasons for attrition.

Intervention/Treatment Fidelity - Introduction

In this section, you will learn more about intervention/treatment fidelity in randomized controlled trials, including:

- Core components of treatment fidelity
- Why treatment fidelity is important
- How to induce treatment fidelity
- How to assess treatment fidelity

Core Components of Treatment Fidelity

As the independent variable, the treatment plays the lead role in a trial testing whether an intervention works. Procedures need to be in place to ensure that the intervention is implemented as intended. This is called establishing treatment fidelity.

- Treatment integrity

  An intervention is based upon a theory of behavior change. Conceptually, the integrity or construct validity of an intervention is the degree to which the treatment protocol operationalizes the influences that the theory posits cause change. Pragmatically, treatment fidelity describes whether the interventionist delivered the treatment as planned.

- Treatment differentiation

  Conceptually, this means that the treatment protocol was operationalized and the interventionists delivered the active change ingredients specified by theory and did not deliver other change elements proscribed by the protocol.

  Differentiation also means that the control condition lacked the active change elements theorized to be integral to the intervention's effectiveness. In designs that test more than one active intervention condition, the theoretically active change ingredients should differ as intended. Distinctive elements of the different treatments should not "bleed," i.e., be implemented in an inappropriate condition.

Why Treatment Fidelity is Important

Treatment fidelity can make or break the test of a behavioral intervention. Here's why.

Treatment fidelity:

- **Preserves internal validity** against:
  - **Type I error,** in which the researcher finds a significant treatment effect that does not really exist. The effect may have come about because the trial lacked treatment fidelity: e.g., an unintended treatment ingredient was added to the intervention
  - **Type II error,** in which case the researcher finds no treatment effect when an effect truly is present. The trial may have been invalid because

it lacked treatment fidelity: i.e., the treatment wasn't actually administered as intended

- **Improves power** (research efficiency) by reducing unintended variability in the treatment effect
- **Allows replication and dissemination** of a well-characterized intervention

Establishing Treatment Fidelity

Several kinds of activities are needed to induce treatment fidelity.

- **Operationalize the Intervention**

  The treatment protocol in an RCT operationalizes a theory of behavior change. The protocol specifies the components, sequence, and underlying rationale for the intervention.

  - Some interventions can be standardized and delivered in a fixed format. Examples include media campaigns, newsletters, or web-based intervention programs.
  - The series of decision rules that underlies a treatment can be conveyed to interventionists as a set of algorithms (i.e., â€œif, thenâ€☐ statements).
  - A treatment protocol can also be written out in a treatment manual
    - A treatment manual describes the content of the series of treatment sessions. It explains the theory underlying the intervention well enough to help interventionists decide how to manage unexpected issues that arise during treatment.
    - Manuals help to standardize the delivery of a treatment and can be disseminated to facilitate the training of interventionists.
- **Train the Interventionists**

  Train interventionists to a prespecified level of competence. Training usually emphasizes both therapeutic expertise and specific intervention content. Stylistic competences may include communication clarity, rapport-building, and clinical skills. Content competencies include mastery of session content and intervention techniques.

- Some very complex psychosocial interventions (e.g., psychotherapies) require therapists who have a high level of professional training and experience, as well as good clinical skills. Other interventions can be performed by less highly trained interventionists or even by peers. The treatment protocol or manual should indicate the requisite educational background.
- Regardless of whether interventionists are professionals or peers, they need training in order to standardize treatment delivery.

- **Supervise the Interventionists**

  Supervise interventionists to ensure that they are delivering an intervention faithfully and without drift over time.

Assessing Fidelity

In addition to making efforts to induce fidelity, an investigator needs to assess whether those efforts were successful. A fidelity checklist is a useful tool to assess treatment fidelity.

- Fidelity checklists specify stylistic and content elements that need to be implemented during the study to preserve treatment integrity.
- A checklist prescribes required treatment deliverables and proscribes elements that represent contamination (blending) from another study condition.

Click here to see an example of a Treatment Fidelity Checklist.

Assessing Fidelity

Fidelity checklists can be used to train study interventionists. Meeting pre-specified performance criteria on a checklist can be taken as evidence that interventionists have achieved competence in the intervention.

Achieving the competency standard may result in certification indicating that an

interventionist is qualified to begin delivering treatment in a trial.

Assessing Fidelity

It is not enough to induce fidelity just at the start of a trial. Efforts are needed to ensure that fidelity remains high throughout the trial. That requires taping or coding live sessions on an ongoing basis using a fidelity checklist.

Preferably, interventionists should remain blind to which of their sessions will be coded. If fidelity falls below a certain criterion, interventionists will need to be retrained to eliminate drift.

III. Conducting an RCT

Conducting an RCT - Introduction

In this section, you will learn more about:

- Recruitment of participants
- Retention of participants
- Adverse events reporting
- Moderation/mediation
- Statistical significance testing

Planning for Recruitment and Retention

Recruitment and retention of participants in RCTs are often challenging. Good planning and close ongoing monitoring can make a difference. Click each item to learn more about steps that investigators can take to increase their odds of recruiting and retaining the desired sample.

- **A Recruitment Plan**

  A good recruitment plan:

  - is detailed
  - includes multiple strategies
  - specifies interim recruitment goals
  - anticipates challenges and potential solutions
  - remains flexible to cope with unanticipated problems
- **Recruitment Sources**

  The recruitment source for an RCT depends on the population of interest. Common recruitment sources include: clinics, private practices, community centers, schools, or media advertisements.

Recruitment

An investigator needs to make several decisions when crafting a recruitment plan.

- **How many study candidates will need to be screened in order to randomize the desired number?**

  Studies with stringent exclusion criteria require a very large pool of potential participants because they screen out so many.

- **Proactive (actively seeking out participants) vs. Reactive (waiting for volunteers to approach the team) recruitment strategies**

>

Provider or colleague referrals, flyers and media advertising fall at the reactive end of the recruitment continuum. Active case finding via review of medical records or random digit dialing represent more proactive approaches. Reactive is ordinarily less expensive than proactive recruitment, but open to question about the representativeness of those who seek out research participation.

- **Will financial incentives be offered?**

  These can range from providing travel, parking, child care, and meals, to paying participants for their time (see http://ohsr.od.nih.gov/info/sheet20.html for federal guidelines on incentives). Incentives make recruitment easier, but their use can also raise questions about the external validity of the trial.

Monitoring and Retention

There are many reasons why participants drop out or fail to complete assessments in an RCT. Some are unavoidable (e.g., death, relocation). There are, however, certain things that investigators can do to retain participants and, thereby, reduce drop-out and missing data. Most important is that study personnel establish a positive relationship with research participants.

- Branding of a study and perceived ownership by participants help to foster retention. Burden and drop-out are reduced by keeping assessments brief and allowing as much flexibility as treatment fidelity can accommodate. Conducting ongoing evaluation of progress toward interim recruitment goals helps to diagnose problems.
- It is critically important that research staff make every effort to complete outcome assessments with all randomized study participants â€" including any who have stopped participating in treatment.

Adverse Events and Serious

Adverse Events

Methods need to be in place to identify and report adverse events that occur during the course of a study.

- An adverse event (AE) is an undesirable health occurrence that occurs during the trial and that may or may not have a causal relationship to the treatment.
- A serious adverse event (SAE) is defined as something life-threatening, requiring or prolonging hospitalization and/or creating significant disability.

For example...

For example...

A suicide attempt would be considered an SAE in a study of any treatment. The SAE needs to be reported regardless of whether it bears any relationship to the treatment or the problem being studied.

Assessing and Reporting Adverse Events

The reason to track adverse events is that they might suggest that there are risks associated with the intervention being studied.

Depending on the severity and frequency of adverse events, investigators and data safety monitors may have to decide to terminate the trial prematurely.

- **Methods for Large RCTs**

  For most large RCTs, a data safety monitoring board (DSMB) (a group of people charged with monitoring participant safety and data quality) should be assembled before the trial begins. SAEs need to be reported immediately to the DSMB, the institutional review board, and the funding agency, which determine if any action needs to be taken. Depending on the severity and frequency of adverse events, the DSMB is empowered to terminate the trial prematurely.

- **Methods for Smaller Trials**

  Smaller trials usually have a data safety monitoring plan (DSMP) rather than a full board. The DSMP may have one or two people or an institutional committee who are designated as data safety monitors.

Moderation and Mediation

Click each item to learn more about testing moderation and mediation.

- **Testing Moderation**

  An investigator may hypothesize that a treatment works better for certain kinds of people or certain kinds of circumstances. For example, does an intervention produce benefits for men but harm for women? Is it more effective when implemented in individualistic than collectivist cultures?

  To test such hypotheses, the investigator pre-specifies the moderator variable of interest before the intervention is delivered. The test of moderation is whether the variable influences the strength or direction of the association between the intervention and outcome.

- **Testing Mediation**

  Researchers often want to test whether their intervention produced the observed study outcome via the change mechanism that they hypothesized. If so, the investigator hypothesizes mechanisms of change (mediators) and measures them over the course of the study. To test mediation hypotheses, the researcher uses statistical methods, such as hierarchical regression analysis,

to:

- Determine that the intervention predicts the outcome
- Establish that both the intervention and the outcome covary as expected with the hypothesized mediating variable
- Establish that controlling for the mediator explains part of the intervention's effect on the outcome

For example...

For example...

An investigator wishes to understand how a parenting intervention improves symptoms of attention deficit hyperactivity disorder (ADHD) in children. The investigator hypothesizes that the intervention achieves its benefit by improving communication and decreasing conflict in the parent-child relationship.

To test mediation, the investigator would examine:

- whether the intervention resulted in positive outcome (e.g. did being randomized to the intervention predict decreased ADHD symptoms?)
- whether the intervention was associated with changes in parent-child communication (e.g. was exposure to the intervention associated with decreased conflict in parent-child relationship?)
- whether the change in the outcome was associated with change in parent-child communication (e.g. were decrease in ADHD symptoms correlated with a decrease in conflict?); and
- whether accounting for the association between conflict and ADHD symptoms weakens the association between the intervention and ADHD symptoms. (That would suggest that decreased conflict is a â€œworkingâ€□ ingredient by which the parenting intervention lessens ADHD symptoms.)

Statistical Significance Testing and Beyond

Click each item to learn more about statistical significance testing and beyond.

- **Estimation vs. Statistical Significance Testing**

  Determination of whether a treatment works has been based traditionally on statistical significance testing. However, whether the effect of a treatment reaches a conventional significance level ($p < .05$) depends heavily on factors such as sample size.

  Recently, scientists have moved towards reporting effect sizes and confidence intervals, as these provide meaningful information about magnitude of change.

- **Assessing the Effect Size of an Intervention**

  An effect size describes the magnitude of an intervention's effect on the study outcome. In the case of RCTs, the effect size represents the magnitude of the difference between the control and intervention conditions on a key outcome variable adjusted for the standard deviation of either group.

  Effect sizes are:

  - necessary to compute a power analysis (needed to determine sample size for a new study and an essential component of most grant applications)
  - increasingly regarded as a required element when reporting the outcome of an intervention study

- **Assessing Clinical Significance**

  When testing interventions that address health problems, it is important to examine whether the observed change is clinically meaningful.

  Some ways that investigators portray the clinical significance of their findings are:

  - The proportion of patients who move from exhibiting clinical or dysfunctional levels of symptoms/behavior to functional levels.
  - The Number Needed to Treat (NNT) expresses the number of patients who need to receive the intervention to produce one good outcome compared to control. NNT is a widely used index of clinical significance.

  For example...

For example...

In a multi-site study involving 5,000 women, an investigator studied the efficacy of an intervention to reduce depression among women with breast cancer. An average pre-post decrease of 5 points on a depression scale was observed. This change was statistically significant, probably due to the very large sample size and power to detect very small effects. However, even though depression scores decreased after the intervention, the average score remained in the clinical range, indicating severe depression. Do you think this intervention was effective?

IV. Data Analysis and Results

Data Analysis and Results

In this section, you will learn more about data analysis in randomized controlled trials, including:

- Data analysis decisions
- Sample size and statistical power
- Data analytic techniques for continuous and categorical outcomes

Issues in Data Analysis

Certain basic decisions need to be resolved before proceeding with data analysis for

an RCT. What data will be analyzed and what analytic approach will be used?

- **Which participants will be analyzed?**
  - Once randomized, all analyzed. That policy, known as the intent to treat (ITT) analysis principle, is a cornerstone of good clinical trial practice. All participants who were randomized and entered the trial need to be included in the analysis in the condition to which they were assigned, regardless of whether they completed the trial, or may even have switched over to receive the incorrect treatment.
  - It may seem odd to include people in the analysis who dropped out before receiving any treatment at all. The rationale is that ITT analysis tests a treatment as policy. The analysis addresses the question of whether the study treatment, if made available to the population, would be superior to an alternative intervention. Offering an intervention that many people would elect not to undergo represents poor public health policy, even if those who did complete the treatment fared very well. For that reason, the disposition of the sample from the moment they learn their allocation is relevant to evaluating the treatment.
  - Per protocol analysis represents the opposite end of the spectrum from ITT analysis. A per protocol approach includes in the analysis only those cases who completed treatment. Results of per protocol analysis represent the best case treatment results that could be achieved if the study sample were retained and remained compliant with treatment.
- **How many analyses?**

  Each of the study's primary aims should test a hypothesis and link to an analytic plan. When planning the trial, statistical power should be computed to choose a sample size adequate to detect the predicted effect, if one is present.

  - Like the number of primary outcomes, the number of aims (hypotheses) should be limited.
  - In an RCT testing whether a treatment works, the main study aim usually tests a predicted difference in the primary outcome between intervention and control at either 1) a final follow-up point or 2) the trajectory of change over time.
  - Investigators should interpret non-primary analyses as post hoc and exploratory, especially if these were not planned in advance. There is great danger of cherry-picking or fishing by performing multiple comparisons and selectively reporting those that turn out as hoped.
  - Exploratory analyses may be undertaken by using a more conservative significance level adjusted for multiple comparisons (e.g., .001 versus .05). However, it should be noted that any findings require replication because they may have arisen by chance alone.
- **Subgroup Analyses**

Whether to perform subgroup analyses is a topic of some controversy.

- **Planned subgroup analyses.** In a few instances, a study may have been designed and powered to test whether a treatment works better for one demographic group (e.g., females) than another (e.g., males). In that case, testing a hypothesized treatment by demographic group interaction would be a primary aim that definitely needs be tested.
- **Exploratory subgroup analyses.** More often, many different treatment-by-subgroup interactions will be explored. Those analyses can support hypothesis generation. The important caveat is the need to remember that unplanned subgroup analyses are done in the context of discovery rather than confirmation. Any findings require replication in another trial.

For example...

For example...

A post hoc analysis of an RCT found an interaction indicating that treating depression after a heart attack decreased the risk of repeat heart attacks for white men, but increased the risk for white women. On the basis of that evidence, an insurance company decides not to pay for depression treatment for women who become depressed after having a heart attack. The company says their decision reflects evidence-based practice. Do you agree?

Sample Size and Statistical Power

Investigators can make two types of errors when testing hypotheses.

- **Type I Error**

  Type I error is usually considered the more serious of the two. This error occurs when findings lead an investigator to reject the null hypothesis of no difference

between treatment and control, when in fact there is no true difference between them.

- **Type II Error**

  Type II error occurs when, based on study findings, the investigator fails to reject the null hypothesis. However, the null hypothesis should have been rejected because there is a real difference between treatment and control.

  To reduce the risk of a Type II error, the investigator should conduct a power analysis before doing the study, in order to determine the sample size needed to detect an effect of the expected size.

Computing Sample Size and Statistical Power

A power calculation can be computed by hand or by computer software. The computations take into account (usually pre-specified) values of:

- a or type I error bound (usually .05 or .01);
- ÃŸ or type II error bound (usually .80 or .90);
- whether the test to be performed is two-tailed or one-tailed. (A two-tailed test is used when a significant difference in any direction is of interest. A one-tailed test may be used when only group differences in a prespecified direction will be reported);
- whether the outcome to be estimated is continuous or categorical;
- estimated effect size (based on pilot data or previous research); and
- planned statistical test.

Power calculations can be conducted using online or downloaded software packages such as G*Power or OpenStat. A helpful website reviewing a variety of free statistical software, including software to conduct power calculations for different types of analyses, is: http://statpages.org/javasta2.html

Data Analytic Techniques

The statistical method used in an RCT is designed to compare the intervention and control conditions' influence on the health outcome of interest to see if the effect differs.

In choosing an analytic technique, the first question that needs to be answered is whether the outcome varies continuously (like symptom severity) or categorically (like being hospitalized or quitting smoking).

Continuous Outcome Variables

In an RCT with a continuous outcome, the study hypothesis usually predicts that treatment groups will differ on the study outcome at final follow-up. Alternatively, the prediction may be that the outcome changes at a different rate for intervention than control (e.g., symptoms decrease more rapidly). Several analytic techniques address this type of question.

- **Analysis of Variance (ANOVA)**

  ANOVA has a long history of use to analyze such data. It examines the effect of an independent categorical factor (such as treatment allocation) on a continuous outcome variable.

- **ANOVA Variants and Extensions**
  - univariate ANOVA
  - repeated measures ANOVA
  - analysis of covariance (ANCOVA)
  - multivariate analysis of variance (MANOVA)
- **Positive Features of ANOVA**

  ANOVA can:

  - address multiple measures
  - handle repeated observations over time
  - account for covariates
  - handle multiple independent and dependent variables
  - use categorical or continuous predictors

- handle fixed factors (where the levels studied comprise the entire population of interest) and random factors (where the levels studied represent a sample of a larger population)

Problems with ANOVA

ANOVA, especially as implemented by major software producers, makes several assumptions that are often unwarranted:

- independence of observations (i.e., that repeated measures are uncorrelated)
- linear and homogeneous slopes of change (i.e., that change over time is linear and parallel across conditions)
- data are missing completely at random (MCAR) (i.e., that there is no systematic pattern to missing data)

Data Analytic Techniques - Alternative Analyses for Continuous Outcomes

Although ANOVA is still used, newer statistical modeling strategies make less restrictive assumptions and have become more frequently used. That is, at least in part, because they make the less restrictive assumption that missing data are missing at random (MAR), rather than missing completely at random (MCAR).

- **Latent Growth Models**

  These models are based on structural equation modeling. They allow for the estimation of individual means, variances, regression coefficients, and covariances for the randomâ€"latent or unobservedâ€"effects that characterize

each individual's growth pattern.

Latent growth models can be used to:

- model individual-level growth trajectories over time
- examine how growth patterns depend on covariates (such as the intervention condition)

- **Growth Mixture Models**

  Growth mixture models were developed to address the fact that intervention effects can vary for categories of people who are launched on different trajectories of change over time. For instance, a new depression treatment might be helpful for patients whose initial symptoms are high and improving. Conversely, the treatment might be ineffective for patients whose symptoms have been high and stable for a long time.

- **Missing Data**

  Decisions about how to handle missing data are of critical importance when using growth modeling approaches.

  There are several valid techniques to deal with missing data, including:

  - Maximum likelihood: models all the data that were obtained, adds some error variance, and estimates the parameter values that make the observed data maximally likely. Then the approach uses the parameter estimates to estimate the missing values. Next it re-estimates the parameters based on the filled-in data, and so forth until the solution stabilizes.
  - Multiple imputations: creates a set of complete data by imputing each missing value using existing values from other variables. The imputation process is repeated many times and each data set is statistically analyzed. Results of these analyses are then combined into one overall analysis.

- **Multi-level Models**

  Multi-level models apply growth curve modeling in a hierarchical structure. Approaches go by many different names and provide an array of modeling techniques:

  - multi-level models
  - mixed effects models
  - hierarchical models
  - random effects models

  The parameters in a multi-level model are allowed to vary at more than one level. For example, multi-level modeling can be used to model change over

time within individuals and between individuals within a single statistical model.

Categorical Outcome Variables

Many outcome variables are categorical rather than continuous. Examples include outcomes like hospitalization, suicide, or taking up smoking. Different analytic techniques are needed to test intervention effects on categorical variables. These approaches include:

- Generalized estimating equations (GEE): used to analyze longitudinal correlated response data when outcomes are binary. For example, compares the effect of treatment versus control on tobacco abstinence over time
- Survival analysis: involves the modeling of time to event data, such as time to death, or time to recovery

Reporting Results - Introduction

In this section, you'll learn about reporting data from randomized controlled trials using CONSORT reporting guidelines.

Guidelines for Reporting Results

The Consolidated Standards of Reporting Trials (CONSORT) has become the gold standard for reporting the results of RCTs.

- The CONSORT guidelines prompt researchers to report sufficient detail about

the trial so that readers can critically appraise it.

- These guidelines are supported by most medical journals, the American Psychological Association, and most behavioral and social science journals that publish RCTs.
- The CONSORT statement can be found at www.consort-statement.org. Additional requirements for reporting behavioral clinical trials can be found in Boutron et al (2008).

CONSORT Guidelines

CONSORT guidelines for reporting behavioral RCTs require investigators to report details such as:

- how participants were allocated to interventions
- scientific background and explanation of rationale for study
- eligibility criteria for interventionists, participants, and the settings and locations where the data were collected
- details about the different intervention conditions, how they were implemented, and how treatment fidelity was enhanced and assessed
- specific objectives and hypotheses
- the primary and secondary measures

Click here to see a CONSORT Flow Diagram.

To see more CONSORT guidelines for reporting behavioral RCTs, visit the CONSORT website at www.consort-statement.org.

V. Ethical Considerations

Ethical Issues

Investigators are responsible to uphold ethical standards and guidelines.

- Declaration of Helsinki. Developed by the World Medical Association, this set of ethical principles guides medical researchers in conducting research on human subjects. It can be found online at: http://www.wma.net/en/30publications/10policies/b3/index.html
- American Psychological Association Ethical Principles for Psychologists and Code of Conduct can be found at: http://www.apa.org/ethics/code2002.html

Ethical Issues

In designing their research, investigators should consider these important ethical guidelines.

- **Competence**

  A researcher's background, experience, and expertise should indicate competence in the research topic area.

- **Stage of Research**

  RCTs are often defined by what phase of research they are in (Phase I, II, II), corresponding to the stage of intervention development. Early in intervention development, potential risks associated with a new treatment may not be known.

- **Research with Special Populations**

  Ethical standards prohibit the exclusion of special populations without a scientifically sound reason. Conversely, special populations should not be studied out of convenience.

  Special populations include:

  - women
  - children

    - prisoners
    - individuals diagnosed with HIV/AIDS
    - depressed and suicidal individuals
- **Selection of Appropriate Comparison Groups**

  Ethical standards must be considered when selecting a comparison group. Considerations include what the current standard of care is, and whether a no-treatment or wait-list control group is ethical.

- **Selection of Assessment Instruments**

  Ethical standards require study instruments to have adequate psychometric properties.

- **Institutional Approval**

  Ethical standards require that any research protocol be approved by an Institutional Review Board (IRB) to ensure the protection of human subjects.

  Some procedures for safeguarding human subjects include:

    - Informed consent procedures
    - Procedures to safeguard confidentiality
    - Protocols to preserve safety and address adverse events
    - Reporting study results
- **Conflict of Interest**

  Ethical principles provide guidance in situations where there are competing interests, values, or commitments. In research, this usually refers to situations in which financial or other personal considerations may compromise an investigator's judgment in the conduct of research or the reporting of results. Conflicts of interest may threaten the integrity of research, erode public trust in science, and negatively affect the rights and welfare of human subjects.

RCTs in Diverse Populations

Ethnic minorities continue to be underrepresented in RCT research. This is especially unfortunate because minorities represent a growing proportion of the U.S. population and have an increased risk for some health problems.

>

- **Inclusion of Ethnic Minorities in RCTs**

  The National Institutes of Health (NIH) mandates that there be adequate representation of minorities in all research that they fund.

  - This policy ensures that interventions developed with federal funding are broadly generalizable to diverse populations

  Investigators are well-advised to expect to encounter various barriers when trying to recruit women and minorities.

  For example...

  However, it is not as simple as including participants; issues such as the cultural relevance of measures and intervention content also need to be addressed.

- **Cultural Relevance of Treatments/Interventions**

  There are many issues that are important to consider before embarking on RCT research with ethnic minority populations:

  - Conceptualizations of mental health or psychopathology within particular cultures should be considered
    - Not all cultures perceive mental health or illness in the same way that Western science does.
  - It is necessary to consider how the intervention being tested was developed
    - Was it initially tested on European Americans?
    - What modifications might need to be incorporated to make it relevant to non-European American populations?
  - Investigators should consider issues such as acculturation or ethnic identification in study participants
    - The level of identification with different cultural values could affect participation and outcomes.
- **Issues to Consider when Conducting RCTs with Ethnic Minorities**

  Cultural issues may affect every important area of an RCT design, including:

  - recruitment
  - retention
  - response to intervention
  - investigator's ability to analyze and interpret results

  It's critical that investigators consider these issues ahead of time and have methods in place to address challenges that might occur in these areas.

Possible methods include:

- Including diverse members as part of the study team; this helps ensure that all important perspectives are represented and considered.
- Community-based participatory research (CBPR), a method in which community members from the population of interest act as research partners and participate in the development, implementation, and interpretation of research findings.

For example...

Mistrust of medical and scientific institutions, or language or cultural barriers can make minority recruitment challenging.

To address these challenges, an investigator might:

- Form partnerships with community organizations that serve diverse populations. Ideally, all partners should be involved from the research planning phase and onward.
- Engage diverse community members on an advisory board that guides the research team on problems of importance to community members, recruitment and retention strategies, etc.

Resources

Resources

Further Reading:

- Boutron, I., Moher, D., Altman, D., Schultz, K., Ravaud, P., for the CONSORT

group (2008). Extending the CONSORT Statement to Randomized Trials of Nonpharmacologic Treatment: Explanation and Elaboration. Annals of Internal Medicine, 148(4), 295-309.

- James AACJ, Soler A, Weatherall RRW. Cognitive behavioural therapy for anxiety disorders in children and adolescents. Cochrane Database of Systematic Reviews 2005, Issue 4. Art. No.: CD004690. DOI: 10.1002/14651858.CD004690.pub2.
- Spring, B. (2007). Evidence-based practice in clinical psychology: What it is, why it matters; what you need to know. Journal of Clinical Psychology, 63(7), 611-631.
- Thorpe, K., Zwarenstein, MD, Oxman, MD, et al (2009). A pragmaticâ€"explanatory continuum indicator summary (PRECIS): a tool to help trial designers. Canadian Medical Association Journal, 180(10), E48-E57.
- Wood, L., Egger M. & Gluud, L, et al (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. British Medical Journal, 336: 601-605.

Glossary

Glossary

- **Algorithm**
  A technique explicating â€œifâ€¦thenâ€□ principles in order to standardize decision-making policy and improve the consistency of care delivery. Treatment algorithms are used to systematize research study protocols, practice guidelines, and clinical decisions.
- **Bias**
  Bias is a systematic distortion of the real, true effect that results from the way a study was conducted. This can lead to invalid conclusions about whether an intervention works. Bias in research can make a treatment look better or worse than it really is. Inferences about validity fall into two primary categories: internal and external.

- **Categorical variable**
  A variable defined by membership in a group, class, or category, rather than by rank or by scores on more continuous scales of measurement.

- **Community-based participatory research (CBPR)**
  Research that is conducted as an equal partnership between traditionally trained "expert" researchers and members of a community. In CBPR projects, the community participates fully in all aspects of the research process.

- **Construct validity**
  The extent to which the study tests underlying constructs as intended.

- **Contamination**
  The process through which knowledge, expectations, or communication about the experimental treatment has unintended influence on the non-experimental condition. Contamination occurs when those in the control condition are unintentionally exposed to aspects of the experimental condition.

- **Continuous variable**
  A variable that can take on an infinite number of values; that is, a variable measured on a continuous scale, as opposed to a categorical variable.

- **Control condition**
  A group of participants in a study that are exposed to the conditions of the experiment that do not involve a treatment or exposure to the independent variable.

- **Counterbalanced design**
  Compares two or more groups who receive the treatment and control conditions in different orders. However, participants are not randomly assigned.

- **Data Safety Monitoring Board (DSMB)**
  A Data and Safety Monitoring Board is an independent group of experts who monitor patient safety and treatment efficacy data while a research study is ongoing.

- **Data Safety Monitoring Plan (DSMP)**
  A Data and Safety Monitoring Plan establishes a protocol to monitor participant safety and data quality. A DSMP is used in smaller trials and may have one or two people or an institutional committee who are designated as data safety monitors.

- **Double-blinding**
  Neither the participants nor the investigator know the participants' treatment assignment.

- **Drift**
  Drift describes changes over time in how study interventionists understand and

implement a treatment. Drift represents a serious departure from treatment fidelity in that the treatment being delivered at the end of a trial differs from that being delivered and tested at the trial's beginning.

- **Effectiveness**
  The capacity of an intervention to produce an effect under usual (or the â€œreal worldâ€□) conditions. Effectiveness trials are usually conducted in community settings, employ local staff as interventionists, and include participants with many comorbid conditions in addition to the target problem.

- **Effect size**
  The magnitude of a treatment effect, independent of sample size. The effect size can be measured as either: a) the standardized difference between the treatment and control group means, or b) the correlation between the treatment group assignment (independent variable) and the outcome (dependent variable).

- **Efficacy**
  The capacity of an intervention to produce an effect under optimal conditions. In health research, efficacy indicates the capacity of a given intervention (e.g. a medicine, medical device, surgical procedure, therapy, or a public health intervention) to produce beneficial change (or therapeutic effect) under favorable conditions. Efficacy trials are often conducted in an academic environment, employ well-trained research staff as interventionists, and exclude participants with comorbid conditions other than the target problem.

- **Endpoint**
  A clinical endpoint refers to occurrence of a disease, symptom, sign or laboratory abnormality that constitutes one of the target outcomes of the trial.

- **Equipoise**
  Clinical equipoise means that there is genuine uncertainty over whether or not the treatment will be beneficial. Even if the researcher truly believes in a hypothesis, there is no actual proof that the benefit exists. Equipoise provides the ethical basis for research that assigns patients to different treatment arms of a clinical trial.

- **Equivalence**
  Parallel versions of the same measure show consistent responses.

- **Evidence consumers**
  Those who consume and use evidence, usually for the purposes of clinical or public health practice, education or teaching.

- **Evidence creators**
  Researchers who conduct studies to produce data that tests a hypothesis or answers a question.

- **Evidence synthesizers**
  Those who acquire, critically appraise, and integrate research findings for the purpose of summarizing evidence regarding a particular question or body of work.

- **Evidence-based**
  Derived from contextualized decision-making that integrates the best available research evidence with consideration of client characteristics (including preferences) and resources.

- **Experiment**
  Describes a situation under which a series of observations are conducted under controlled conditions to study a relationship with the purpose of drawing causal inferences about the relationship. Experiments involve the manipulation of an independent variable, the exposure of different groups of participants to one or more of the conditions being studied, and the measurement of a dependent variable.

- **External validity**
  External validity is the extent to which the results can be generalized to a population of interest. The population is usually defined as the people the intervention is intended to help.

- **Fidelity checklists**
  Specify stylistic and content elements that need to be implemented during treatment to preserve treatment integrity.

- **Fixed Allocation Randomization**
  Each participant has an equal probability of being assigned to either treatment or control and the probability remains constant over the course of the study. That can be achieved by using a table of random digits or randomization software (in SAS, SPSS, and other major software programs).

- **Fixed Block Size**
  In randomized block design, participants are first classified into groups (blocks) of a fixed length (usually 4, 6, or 8), on the basis of a variable that the experimenter wishes to control. Individuals within each block are then randomly assigned to one of several treatment groups.

- **Generalizable**
  The accuracy with which results or findings can be transferred to situations or people other than those originally studied.

- **Generalized estimating equations (GEE)**
  Used to analyze longitudinal correlated response data when outcomes are binary.

- **Intent to Treat (ITT)**

Once randomized, all analyzed.

- **Intermediate endpoint**
  Sometimes used to designate an outcome that is correlated with but not identical to a clinical endpoint.

- **Internal consistency**
  Responses to questionnaire items that measure the same construct are highly intercorrelated.

- **Internal validity**
  Internal validity is the extent to which the results of a study are true. That is, the intervention really did cause the change in behavior. The change was not the result of some other extraneous factor, such as differences in assessment procedures between intervention and control participants.

- **Maximum likelihood**
  Models all the data that were obtained, adds some error variance, and estimates the parameter values that make the observed data maximally likely.

- **Mediator**
  In statistics, a variable that helps to account for the association between an independent and a dependent variable. A mediation model seeks to explicate the mechanism that underlies an observed relationship between an independent variable and a dependent variable via the inclusion of a third explanatory variable, the mediator variable. Rather than hypothesizing a direct causal relationship between the independent variable and the dependent variable, a mediational model hypothesizes that the independent variable causes the mediator variable, which in turn causes the dependent variable.

- **Missing at Random (MAR)**
  Missing at random (MAR) is the alternative of MCAR, suggesting that what caused the data to be missing does not depend upon the missing data itself.

- **Missing Completely at Random (MCAR)**
  MCAR describes the assumption that data are missing completely at random. MCAR assumes that, at any time point, a missing subject or missing data point, occurs for completely random reasons.

- **Moderator**
  In statistics, a variable that alters the direction or strength of the association between other variables. For example, if gender moderates their relationship, two variables may be positively associated among women, but negatively correlated among men.

- **Multiple imputations**
  Creates a set of complete data by imputing each missing value using existing

values from other variables.

- **Number Needed to Treat (NNT)**
  Expresses the number of patients who need to receive the intervention in order to prevent one additional bad outcome.

- **Per protocol analysis**
  Includes in the analysis only those cases who completed treatment.

- **Placebo**
  An inert substance. In a placebo-controlled trial, the group randomized to a placebo arm receives an inert substance that is indistinguishable in appearance from the active drug under investigation. Inclusion of a placebo arm holds expectations about treatment benefit and adherence constant across the control and intervention groups, while allowing a test of the actual pharmacological effects of the active drug.

- **Random assignment**
  To assign participants or other sampling units to the conditions of an experiment at random, that is, in such a way that each participant or sampling unit has an equal chance of being assigned to any particular condition.

- **Randomly permuted blocks**
  Blocks of patients are created such that balance is enforced within each block. For instance, let E stand for experimental group and C for control group, then a block of 4 patients may be assigned to one of EECC, ECEC, ECCE, CEEC, CECE, and CCEE, with equal probabilities of 1/6 each. In each block, there are equal numbers of patients assigned to the experimental and the control group.

- **Selection bias**
  A systematic distortion of evidence that arises because people with certain important characteristics were disproportionately more likely to wind up in one condition. Although random assignment theoretically eliminates selection biases, a bias can still occur.

- **Sensitivity**
  The proportion of people with a condition that the measure correctly identifies.

- **Specificity**
  The proportion of people without a condition that the measure correctly classifies.

- **Stratified randomization**
  A technique in which a population is divided into subgroups (strata) and individuals or cases from each strata are randomly assigned to conditions.

- **Statistical Conclusion Validity**
  The validity of inferences about covariation between two variables.

- **Survival analysis**
  Involves the modeling of time to event data, such as time to death, or time to recovery.

- **Temporal stability**
  A person scores consistently across assessments at two different time points.

- **Time series design**
  Measurements taken before and after an intervention, no control group.

- **Treatment condition**
  The specific intervention condition to which a group or individual is exposed in a research study. For example, in a design employing four groups, each of which is exposed to a different number of sessions of a particular treatment, each â€œdosageâ€ (number of sessions) represents a level of the treatment factor.

- **Treatment differentiation**
  Treatment protocol was operationalized and the interventionists delivered the active change ingredients specified by theory and did not deliver other change elements proscribed by the protocol. Differentiation also means that the control condition lacked the active change elements theorized to be integral to the interventionâ€™s effectiveness. In designs that test more than one active intervention condition, the theoretically active change ingredients should differ as intended. Distinctive elements of the different treatments should not â€œbleed,â€ i.e., be implemented in an inappropriate condition.

- **Treatment fidelity**
  How accurately or faithfully a program (or intervention) is reproduced from a manual, protocol or model. Fidelity is usually measured using a checklist, which is completed by trained raters.

- **Treatment integrity**
  The integrity or construct validity of an intervention is the degree to which the treatment protocol operationalizes the influences that the theory posits cause change. Pragmatically, treatment fidelity describes whether the interventionist delivered the treatment as planned.

- **Treatment manual**
  Provides specific operational guidelines to deliver an intervention. Dissemination and use of a manual maximizes the probability of treatment being conducted consistently across settings, therapists, and clients.

- **Validity**
  Describes how well a test measures what it is intended to measure.

- **Variance**
  A measure of the spread, or dispersion, of scores within a sample. A small

variance indicates highly similar scores, close to the sample mean. A large variance indicates more scores at a distance from the mean and possibly spread over a larger range.