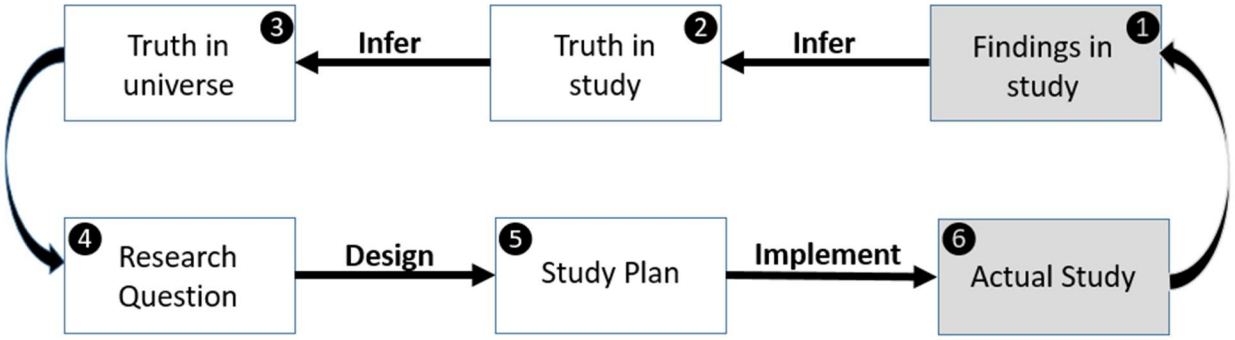
Session 2: Research Design/Methods

Introduction

When designing research, it is important acknowledge what you are actually doing in a research study. It is easy to think that a research study measures what happens in the universe, but this is not accurate. A research study only measures what happens within the confines of the study; the study is only a sample of the universe that we hope accurately represents the universe we hope to learn about. This concept and its associated limitations of research are nicely described by Hulley and colleagues (see Figure).[1]



Adapted from: Hulley, SB. Designing Clinical Research

When we obtain research study results (❶), we use them to make inferences about the truth within the study (❷). Remember that our results are not necessarily truth within the study for a couple reasons. If we were to repeat the same measurements on the same subjects, we might get slightly different results. Also, we will use statistics to say only with limited certainty (usually 95% certainty) that our findings represent what they indicate. We then use what we think is truth in the study to make inferences about truth in the universe represented by our sample (❸). Designing research is essentially the reverse process, where we start with a question about a truth in the universe (❹). We then design a study plan (❺) that we expect will answer that question. Finally, we implement that plan to conduct the actual study (❻) and produce study findings (❶).

Two things should be clear after thinking about this process. First, the study findings we use to infer universal truths are completely dependent on the microcosm we construct to answer our research question; the closer the microcosm is to the universe we are studying, the more accurate our inferences will be. Second, any error or bias we introduce in the design or implementation stages will be carried through, and perhaps amplified, through the conduct of the study and back through the inferences about universal truth. It is therefore critical that, as we design research, we strive to minimize errors and biases through sound research design, and that we are cognizant of the conditions and limitations that our design choices produce.

With that background in mind, the following steps are intended as a guide to help think through some of these issues as you design your study.

Step 1: Comparison Groups

If your research intends to compare differences between groups, how will you test your hypotheses? You must decide what conditions are necessary to be able to compare one condition against one or more others

[1] Hulley, SB et al. 2007. Designing Clinical Research, 3rd edition. Philadelphia, PA: Lippincott, Williams, & Wilkins.

and see if there is a difference. Do/will these conditions exist naturally? Must they be induced by an intervention?

A. Experimental condition(s)
Experimental conditions are typically the "new," "different," "abnormal," or "intervention" conditions. What condition(s) must exist for you to test your hypothesis?

B. Control condition(s)
Control conditions are typically the "usual," "standard," "normal," "healthy," "sham intervention," or "non-intervention" conditions. What conditions must exist to compare against the experimental condition(s) in order to conclude that the unique feature(s) of the experimental condition(s) account for the difference between the groups? Ideally, the control condition differs from the experimental condition in only the predictive factor you are testing. In reality, that can be hard to do. Here are a couple high-yield considerations:

i. *Subjects serving as their own control*: A helpful design is to have every subject serve as their own control, rather than having two separate unrelated groups; such a design increases your ability to detect differences by minimizing random incidental differences between subjects. Self-controls can be accomplished with a design that allows pre/post intervention measurements, or a cross-over design where every subject experiences both the experimental and control interventions. Cross-over designs are not always feasible due to the length of time required for an intervention to be effective, duration/permanency of the intervention effect, or irreversible learning from an initial exposure to the intervention.

ii. *Intervention controls*: In interventional studies, it is best to have the control condition be as similar as possible to the experimental condition, but lacking the component(s) expected to have the effect. For example, rather than comparing acupuncture to no acupuncture, it is preferable to compare acupuncture to a sham acupuncture procedure (where the needles do not puncture the skin…or do puncture the skin, but in non-therapeutic locations).

Step 2: Identify your outcome measures

How will you be able to tell if there is a difference between the experimental and control conditions? The answer to this question will usually be evident from your background/aims/hypotheses, but sometimes you will need to be more specific here. How will you define "differences" between populations or "success" of an intervention? OK, then how will you measure that? What data sources, tests, assays or measures exist that could assess differences in your conditions? Whichever types of measures you choose, they must be meaningful in the field (medically, scientifically, and/or practically), so that someone will want to take action based on your study results.

A. "Hard" vs. surrogate endpoints
"Hard" endpoints are conditions that patients "feel" or experience: morbidities, mortality, success at a goal. Surrogate endpoints are conditions that reflect, or correlate with, meaningful hard outcomes for the condition being studied. It is OK to design your study using either hard or surrogate endpoints, but if

[1] Hulley, SB et al. 2007. Designing Clinical Research, 3rd edition. Philadelphia, PA: Lippincott, Williams, & Wilkins.

a surrogate endpoint is used, make sure it is accepted in the scientific community as having a strong enough correlation with meaningful hard outcome.

B. Subjective vs. Objective endpoints
Subjective endpoints (e.g., clinical diagnosis, subject-reported outcomes, questionnaire responses, observer notes) have greater limitations than objective endpoints (e.g., assay/device-measured outcomes). Objective measures should be maximized; subjective measures are still frequently used and are often unavoidable. If subjective endpoints are used, think of how to minimize potential biases (see Step 6).

C. Quantitative endpoints
Whether subjective or objective, it can be helpful to make your endpoint(s) quantitative, when possible, as opposed to merely descriptive. Quantitative variables will lend themselves better to traditional statistical methods. Objective measures tend to be quantitative by nature; subjective measures can be made quantitative by using, as an example, likert, visual analogue scales, or otherwise categorizing in an ordinal (ordered) manner. Separate statistical methods must be used for purely qualitative research studies that use only qualitative (non-quantitative) analysis.

D. Measurement precision and accuracy
The inherent variability (precision) of measures and their reflection of truth (accuracy) are major factors to consider when choosing outcome measures. You may have the choice between more precise and accurate measures and measures that are less precise and accurate. There is often a trade-off (often cost); the smaller the difference between conditions that needs to be detected, the more important precision becomes. Accuracy is more important when the goal is to evaluate the true value of a test than when the goal is only to compare differences between conditions. Step 6 A and B below further discuss optimizing precision and accuracy.

E. Validated vs. non-validated survey tools
If questionnaires or survey instruments are utilized, it is best to choose a tool that has been validated, through prior research, to produce results that are meaningful in terms of what you intend to measure. This is especially relevant if your measure is intended as a surrogate for a hard condition (e.g., quality of life, future academic success). Validation is less of an issue when survey tools are being used to assess attitudes, opinions, or breadth of opinions, as opposed to outcomes. Survey validation is a form of ensuring accuracy of the measure.

Step 3: Decide on a study design

Most people have a good idea of how they want to approach studying a particular research question. There is a temptation to pursue prospective, randomized study designs because of the rigor they provide in answering questions. However, there are several factors to consider in choosing the optimal study design not only for the research question, but also for the available time and resources available for the project. A rigorous prospective study might be the best option to answer a question definitively, but may not actually be feasible for the investigator to carry out. Investigating rare conditions may be extremely difficult to study prospectively, and alternate designs should be considered. Here are some design features to consider as you plan your research.

[1] Hulley, SB et al. 2007. Designing Clinical Research, 3rd edition. Philadelphia, PA: Lippincott, Williams, & Wilkins.

A. Retrospective vs. Prospective

Prospective designs involve identifying subject and then following them forward in time to test the hypothesis. Retrospective designs involve identifying subject groups that have already been formed and assessing what has already happened to them to test the hypothesis. Prospective design is generally favored over retrospective design because of the ability to control the measures that are collected as opposed to having to rely on data that are already fully in existence (and may not capture everything you need). That said, retrospective studies have the advantage of producing results more quickly and are sometimes the only feasible option. For example, if you are studying a rare disease, it could take decades to encounter enough subjects to study prospectively, but a retrospective study could instantly generate the necessary sample size.

Note that in retrospective studies, all data needed to test the hypothesis are already in existence when the study is initiated. If any data do not yet exist at the time the study has begun, it is not a purely retrospective study (and has, at minimum, a prospective component). This is an important distinction when it comes to human subject protection concerns (which will be discussed in session 4); components involving prospective data collection require different considerations for the informed consent of subjects.

B. Different approaches to prospective study design: interventional vs. observational

i. *Interventional design*; interventional designs involve exposing subjects to a manipulation to observe the effect of the manipulation on them. Observational designs follow the natural course of various conditions to see how they differ. Intervention-based design is generally favored over observational design because of the abilities to more precisely test a hypothesis and to control for more confounding variables; this is especially true when the subjects are randomized to different interventions. Observational studies are limited in examining the effectiveness of treatments or interventions; because they were selected by factors other than chance, it is sometimes very difficult to conclude independence of the treatments from other confounding factors. However, interventional design is not always the best option. Perhaps it is not yet known which types of

[1] Hulley, SB et al. 2007. Designing Clinical Research, 3rd edition. Philadelphia, PA: Lippincott, Williams, & Wilkins.

interventions might work and observation of naturally occurring differences might be needed to generate hypotheses about creating an effective intervention. Perhaps it is not ethical to randomize to two different interventions, and it is necessary to allow each subject to choose his/her own natural course.

C. Different approaches to retrospective study design
   i. *Retrospective Cohort*; this design defines comparison groups (cohorts) of subjects based on differing baseline characteristics or treatment pursued at a point in time in the past. Data are then collected from a later point in time (but still in the past) to examine how outcomes differed. This type of retrospective design is most useful in examining how best to treat or ameliorate a condition.
   ii. *Case-control*; this design defines comparison groups based on the ultimate outcome of interest – cases (individuals in whom the outcome of interest occurred) and controls (individuals in whom the outcome of interest did not occur). Data are then collected about certain characteristics of those individuals, often from an earlier time point, to look for differences between groups. This type of retrospective design is significantly limited by confounders that are difficult to control, but may be particularly useful in the early stages of understanding a condition and searching for possible candidate predictors of the outcome.

Step 4: Identify and address potential confounders

Besides the factor(s) you hypothesize to cause the difference between conditions, what else could affect any potential differences in your comparison groups (positive or negative effects)? For this, you need to take a step back from your thoughts/hypotheses and play devil's advocate. Be skeptical (cynical?)…be a nay-sayer…and try to pick apart your hypothesis. "I wouldn't believe a positive result because of X…." **This is probably one of the most crucial (yet overlooked and/or under-executed) steps in designing research.** You need to be able to question, critique, and poke holes in an idea you have been incubating and investing yourself in. It is difficult. It can be helpful to bounce your experimental condition(s) and control condition(s) off of a colleague who is not as close to the ideas as you are; because of their distance from the project, sometimes they can see problems you might not have considered.

List as many reasonable factors that could impact your results as possible. This list can be generated by asking two questions: 1) If my hypothesis is confirmed, what else could be causing the difference (what alternative explanations are there)? 2) If my hypothesis is not supported, what factors could be influencing the outcome and preventing a difference from being seen? These factors then need to be controlled for in your experimental design by one of the following:

A. Control for the factor

   One of the best way to mitigate a potential influential factor is to ensure the factor is represented in the control conditions (or including additional control conditions to account for the factor). Consider this example: you would like to test the effect of a behavioral intervention on depression scores, and you will randomize subjects to receive the intervention compared to no intervention. But how will you know that the intervention was what resulted in any detected difference between groups? Perhaps, just spending

[1] Hulley, SB et al. 2007. Designing Clinical Research, 3rd edition. Philadelphia, PA: Lippincott, Williams, & Wilkins.

the extra time with the subjects in the experimental group led to differences. This confounding factor could be controlled for by spending equal time visiting with subjects in the control group. Setting up adequate control conditions sounds straightforward, but it is surprising how often experiments are inadequately controlled. Hopefully, you already thought of how to isolate the causative factors when coming up with your comparison groups. Thinking critically about this again now may help you go back and redefine or refine your comparison groups. Incorporating factors into the control condition is more intuitive when the factors are study procedures. Sometimes, however, it is subject characteristics that may be confounders; while not as easy, these characteristics can be balanced in the control condition through either stratification or matching.

i. *Stratification*. If you are randomizing subjects to different conditions, you can stratify the randomization such that the potential confounder is likely to be randomly balanced between groups and therefore controlled for. For example, let's say you want to study the effects of a drug vs. placebo to restore bowel function following a surgical procedure which could either be performed as open surgery or endoscopically. These two surgical approaches have different impacts on resumption of bowel function, but it is important to you to include both procedure types in this study (i.e., you are unwilling to narrow the population to just one or the other). It would not be appropriate to randomize subjects to a particular procedure type; treatment approach is something chosen as part of clinical care. If you randomize subjects and there ends up being an imbalance in surgical approaches, the study results may be driven entirely by procedure type and therefore may not be valid. One possible solution is to stratify the randomization of subjects with respect to procedure type. Essentially, all open surgical subjects are randomized to study drug or placebo, and all endoscopic subjects are separately randomized to study drug or placebo. This way, your final analysis population will have roughly equal numbers of each procedure type and this factor will not impact comparison across arms. One limitation of stratification is that only a small number of variables (usually one or two) can be controlled by stratification because each additional stratum subdivides the study population into increasingly smaller subgroups.

ii. *Matching*. If you have at least two comparison groups, and you know that there is a factor that can influence your outcome, it is possible to match subjects in the different comparison groups with respect to that variable. A common example would be age and gender having independent effects on your outcome measure, such as muscle mass. An imbalance in males vs. females, or older vs. younger ages in your groups could influence the study results. Subjects in one group (typically the control group) can be selected to be age and gender-matched with subjects already identified in the experimental group. As opposed to stratification, matching is used when groups are not randomized (i.e., when there is a cohort design). Matching is most effective when you have one relatively narrow group (e.g., the disease/condition of interest) and are comparing to a relatively broad group (e.g., healthy controls).

B. <u>Narrow the study population</u>

It is sometimes possible to eliminate a factor that could influence or skew your outcome by through inclusion/exclusion criteria. For example, if you are concerned that individuals who regularly exercise might not respond as well to your exercise training program as sedentary individuals, you can exclude

[1] Hulley, SB et al. 2007. Designing Clinical Research, 3rd edition. Philadelphia, PA: Lippincott, Williams, & Wilkins.

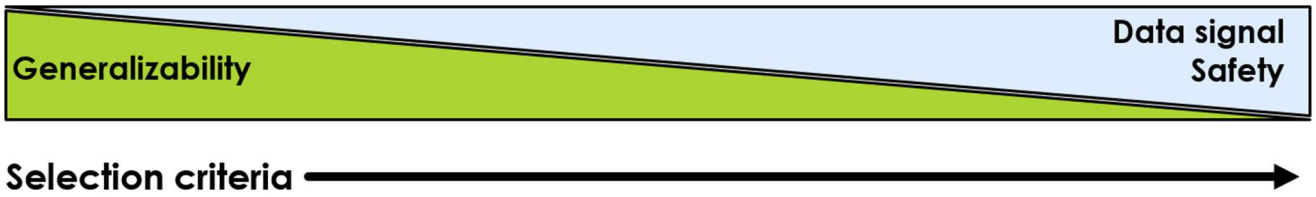
anyone who exercises more than two times per week from the study sample. See Step 5 for more detail on this approach.

C. Record the factors as study data so that they can be controlled for in statistical analysis
If you cannot control for a variable through study design or narrowing your population, you may need to adjust for it in analysis. Controlling for, or assessing the impact of, variables during analysis can be difficult. If this is something you think you might need to do, you should discuss this with a statistician.

Any confounder that you are unable to minimize in some way will need to be considered as a limitation of your study. Even if minimized by one of the methods above, some confounders are still limitations when it comes to study conclusions.

Step 5: Define Subject Populations

It is critical to define the characteristics of the subjects that will be studied. Remember that you are selecting a sample of a larger population to study; you will then use your study results to make conclusions regarding the larger population. Therefore, there is a natural tension between scientific rigor and generalizability to the larger population. The more highly-selected your research subjects, the "cleaner" your data signal may be because potential confounders have been minimized, but the less generalizable your results may be to the larger population because you have selected a narrower and narrower slice of the population. In early-phase research (e.g., pilot studies), it is often a favorable trade-off to sacrifice generalizability to improve the ability to detect a scientific signal.



Subject selection is also important when it comes to protection of human subjects (scrutinized by IRBs and regulatory authorities). Some research procedures you would like to perform may be riskier in individuals with certain characteristics than in other individuals. As we will cover in future sessions, it is your responsibility to ensure that risks to research subjects are minimized. Minimizing risks may involve excluding some individuals that are at higher risk even though they are part of the population at large. Again, this creates tension between safety and generalizability, and it is a reality of conducting human research.

Given the concepts described above, it is a good exercise to think of selection criteria separately based on whether they are for scientific rigor (data quality) or for subject safety. With this in mind, let's consider subject selection criteria, which are traditionally divided into **inclusion criteria** and **exclusion criteria**.

A. ***Inclusion Criteria*** are the characteristics that potential subjects must exhibit to be enrolled in the research. These are the factors that narrow the scope of your investigation from the world's population

[1] Hulley, SB et al. 2007. Designing Clinical Research, 3rd edition. Philadelphia, PA: Lippincott, Williams, & Wilkins.

down to the disease or condition of interest. Rather than a disease or condition, you might be studying normal physiology and want to select subjects that are healthy and free of significant disease. If you are recruiting different baseline cohorts, you will need to define these cohorts separately (i.e., separate inclusion criteria for diseased cohort and healthy control cohort). A study typically has fewer inclusion criteria than exclusion criteria. Inclusion criteria should:

i. include one or more criteria that will describe/define the disease state or condition you are studying (if applicable). After reading the inclusion criteria, it should be obvious to a reader how you are identifying the kinds of people that will help answer the question (this will then feed directly into subject recruitment).
ii. almost always be positively-stated. Note that it is easy to frame exclusion criteria as inclusion criteria ("not pregnant"); in general, it is cleanest to keep attributes that must be present as inclusion criteria and attributes that must be absent as exclusion criteria, and try to minimize any redundancy.
iii. be kept to the minimum necessary to allow for a precise and safe study while maintaining as much generalizability as possible. Make sure your inclusion criteria are justified by either promoting cleaner data or subject safety necessary before listing them.
iv. almost always include age range, given the importance of age in defining disease and safety in being subjected to procedures. Try to always include an upper age range (many IRBs *require* listing an upper limit); in other words, there is typically an age at which your study procedures would not be safe and the age should be capped.
v. strive to maximize data quality by limiting potential confounders. If you have identified a potential confounding factor (see Step 4) that is best controlled by limiting the types of individuals enrolled, it may be addressed through an inclusion criterion. For example, if you want to study normal respiratory epithelium, you may decide to enroll only individuals who have never smoked tobacco.
vi. strive to minimize subject risk by limiting the population to subjects that have the lowest risk for the procedures you wish to use. For example, if you will have subjects undergo a series of maximal exercise tests to exhaustion, you may select individuals who are active and exercise regularly at a level close to what you will have them do in the study.

B. ***Exclusion Criteria*** are the characteristics that potential subjects must not exhibit to be enrolled in the research. These are the factors that further fine-tune the population defined by your inclusion criteria to further promote data quality and patient safety. Exclusion criteria should:
   i. be kept to as few as possible to allow a precise and safe study. Make sure your exclusion criteria are justified, by either promoting cleaner data or subject safety, before listing them.
   ii. strive to maximize data quality by eliminating potential confounders. If you have identified a potential confounding factor (see Step 4) that is best controlled by limiting the types of individuals enrolled, it may be addressed through an exclusion criterion. For example, if you want to study the effect of diet on inflammatory markers, you may decide to exclude any individuals with a recent infection or other acute inflammatory condition.

[1] Hulley, SB et al. 2007. Designing Clinical Research, 3rd edition. Philadelphia, PA: Lippincott, Williams, & Wilkins.

iii. strive as to minimize subject risk by excluding potential subjects that have a higher risk of undergoing the procedures you wish to use. For example, if you know that all subjects will be randomized in one arm to aspirin therapy, excluding individuals with aspirin allergy would be important.

Step 6: Minimizing Variability and Bias

You have already thought about your outcome measures (see Step 2). Each measurement will have a certain inherent degree of variability; the more variability in your measurement, the less power your study will have to detect differences in your outcome for the same number of subjects. How can you improve the precision of your measurements and increase your study power (or decrease the sample size needed)? You have also thought about potential confounding factors (see Step 4). In doing so, you may have identified potential confounding factors that you are unable or unwilling to control through limiting the study population; these confounders can introduce bias into your results if not controlled for. In addition to confounders, have you considered any potentially subjective influences that could bias your measurements or results? For example, if you are rating the percent body-surface area of a subject's rash and you know whether the subject received active treatment or placebo, how will you ensure that the knowledge of treatment assignment will not subconsciously bias your assessment? **Considering variability and bias is probably the second-most under-executed step in designing research.** Let's explore some methods to minimize bias from uncontrolled confounders and subjectivity.

A. ***Measurement Precision.*** All measures have some degree of natural variability around the "true" result. Precision is reflected by the variability in the test result when measuring the same standard repeatedly. Many mechanical devices and assays have documentation of precision. For your primary outcome, it is a good idea to know the precision of the measure you are using; it is one of the variables needed to calculate power. The less the precision (and the less the expected difference between groups), the less power a study will have, and the greater number of subjects will be needed to detect a difference. Precision can be optimized through various means:
   i. *Use a consistent measurement procedure.* Performing the measure the same way each time can minimize the amount of random variability translated to the result. Having written instructions for conduct of the measurement and can help accomplish this goal.
   ii. *Training individuals performing measures.* Training each individual who will perform measurements on a consistent, standard technique will minimize inter-observer variability much more than if the individuals learn on their own and develop their own unique "habits" in performing the measurements.
   iii. *Repetition.* With measures that have an inherent random variability, repeating the measurement and taking the average of the values obtained will result in a less variable valuable over time. This is the same concept as sampling a population mean multiple times; the variability in sample means is less than the variability of individual samples.

B. ***Measurement Accuracy.*** All measures are a representation of the "true" result. Accuracy is reflected by how close the measured result is to the true value (is it off by 0.1% or by 30%)?

[1] Hulley, SB et al. 2007. Designing Clinical Research, 3rd edition. Philadelphia, PA: Lippincott, Williams, & Wilkins.

i. *Use a consistent measurement procedure.* As discussed above for precision, this method can also improve accuracy.
ii. *Training individuals performing measures.* As discussed above for precision, this method can also improve accuracy.
iii. *Calibration.* By measuring a known standard value and adjusting the test results to match the known standards, will help the test better reflect reality. Note that this can even be attempted with more subjective measures like questionnaires by validating the questionnaires in certain populations; validation affirms that the result of the tool is meaningful in terms of real outcomes.

C. ***Subjectivity.*** Subjectivity in outcome measurement is a very serious problem in research. At best, subjectivity introduces variability (noise) into the data when different observers all have slightly different views in judging the same outcome; this is common with patient-reported outcomes (where each subject is an independent judge), but also in multicenter studies (where each investigator is an independent judge). At worst, subjectivity results in systematic bias. No matter how objective we think we are, there are subconscious influences that can color our judgement of study outcomes.

Think carefully about your outcome measures. Is there any subjectivity to them? Even seemingly objective endpoints can actually be subjective when scrutinized. Take myocardial infarction, for example; objective, right? The problem is that myocardial infarction includes the need for a clinical presentation of myocardial ischemia which is subject to interpretation of patient symptoms; that interpretation can differ between sites and investigators. Different centers might have different cardiac biomarker assays, or even different interpretations of how high the biomarkers must be before there was a "true" infarction. Subjectivity can be even greater for other clinical diagnoses. If there is a subjective component to any of your outcomes, you may wish to employ one of the following methods to minimize variability or bias.

i. *Blinding.* Perhaps most important of any method is blinding. **Every study should consider some form of blinding**, as blinding can mitigate bias from both investigators and subjects. Blinding is best when both investigators and subjects are blinded to intervention assignment or other predictive variable. The effects of blinding are most obvious in intervention trials, but even in retrospective chart reviews blinding to the outcome variables when collecting and recording the predictor variables can reduce bias. If blinding is utilized, there must be at least one "unblinded" operator on a study; it is OK if the principal investigator is unblinded, as long as everyone interacting with subjects or performing outcome measurements is blinded. Sometimes blinding is not possible due to the nature of the intervention.

ii. *Placebo/Sham control.* It is known that the simple act of providing a treatment (placebo effect) or spending time with a person (Hawthorne effect) can have a positive impact on their health, sense of well-being, or his/her behavior. Therefore, it is best, if not essential, to design control conditions that are "inactive" or "non-therapeutic" versions of the intervention being tested. For drugs, this is usually the placebo ("sugar pill," or other pharmacologically inactive substance). For a device, it may be the same device but inactivated. For an injection or procedure, it is typically a sham procedure (similar to the procedure being tested but without the effective component). If the intervention is some type of counseling or behavioral modification, an effective control would be

[1] Hulley, SB et al. 2007. Designing Clinical Research, 3rd edition. Philadelphia, PA: Lippincott, Williams, & Wilkins.

social visits of the same length of time. Placebo/sham control goes hand in hand with blinding; investigators and subjects should generally be blinded to their treatment assignment to maintain the effect.

iii. *Definitions.* If you are measuring outcomes with subjective components, it is best practice to have prospective definitions for your outcomes in writing. The definitions can be distributed to all investigators performing outcome assessments to minimize the variability in perspectives.

iv. *Training.* Just as with increasing precision in measurements, training investigators on proper approach to interpreting outcome measures can help decrease variability in perspective between investigators/sites. Training usually goes hand in hand with some sort of written definition or procedure to follow.

v. *Adjudication.* A powerful method to limit variability and subjectivity is adjudication, where one or a small group of individuals make the judgment of all outcome measures in the study. This is usually accomplished by core labs or adjudication committees – centrally-located groups that are sent the information about outcome measures to make a central determination of the outcome. Core labs might perform central reads of EKGs, radiographs, or histology. Adjudication committees might review medical records to determine whether the definition of a clinical outcome event (e.g., myocardial infarction) has been met. It is not necessary to have a core lab or a committee; even a single or dual independent read of a film or pathology slide can be sufficient. **Individuals adjudicating outcomes should be blinded to intervention assignment or cohort allocation.**

Step 7: Writing the Procedures Section

Now that you have refined your study design and outcome measures, it is time to write the Procedures section of your protocol. As discussed in Session 1, the procedures are a main focus of IRBs and review committees. It is through your description of study procedures that important assessments such as risks and benefit will be assessed. Therefore, it is important that enough detail is provided so that a reviewer can visualize what a subject will experience. Writing a protocol is a balance between sufficiency and efficiency; too much detail (for example, how blood pressure will be measured, how a subject's arm will be cleansed for a procedure, or the specific MRI protocol used) can detract from the message that needs to reach your audience. Some of those finer details may be important for subject safety, but they can be left for the human subject protections portions of the protocol (Session 4).

In general, the protocol should be a narrative of what subjects will do as they flow through the study. It is best to start by creating a visual diagram (and/or table) of the study procedures and timeline. This will help you organize the procedures section and is a great tool to insert into the protocol to give your audience the same structured overview. The narrative should be visit-focused (i.e., separated by visit or data collection time point), much like the diagram/table showing the study timeline.

Obviously, the approach to the procedures section will be different for a retrospective study where subjects are not going through any study flow. See the brief description at the end of this section.

[1] Hulley, SB et al. 2007. Designing Clinical Research, 3rd edition. Philadelphia, PA: Lippincott, Williams, & Wilkins.

There is no set formula, but here is a suggested flow to your procedures section:

A. ***Outcomes.*** Many protocol templates have a separate section for outcomes that comes before the Procedures section. If yours does not, create one, or list your outcomes at the beginning of the Procedures section. A reviewer should know the outcomes before reading the procedures so that s/he can verify that the procedures are justified and sufficient to assess the outcome(s). For large medical trials regulated by health authorities, outcomes are split into primary outcome (the most important outcome on which the study has been powered) and secondary outcomes. The general rules in these large medical trials are that if the primary outcome is negative, the secondary outcomes are not analyzed; secondary outcomes are analyzed in order, and analysis moves to the next secondary outcome only when the previous one was positive. This model makes sense when all of the outcomes are directly related to the same question. Many academic studies are examining several more loosely-related (or distinct) outcomes, and you may wish to not specify primary vs. secondary as you intend to analyze all. All outcomes should be consistent with, and justified by, your study aims.

B. ***Screening Visit.*** The <u>first</u> thing you will need to do with subjects is obtain their informed consent to participate in the study. This task is often lumped together with some sort of assessment of a subject's eligibility, and usually labeled as the Screening Visit. Note that you cannot perform any procedures (interview, questionnaires, blood draws, etc.) to assess eligibility until a subject has provided informed consent; however, there is a certain amount of "pre-screening" that can be done by examining the medical record and can make subject eligibility assessment much more efficient. Each institution sets its rules for what type of pre-screening that can be done (according to its interpretation of the HIPAA privacy rule), so it is important that you know what is allowed at your institution with regard to pre-screening. It is important that you include the following in this section:
    i. Any narrative needed to provide context or flow of the visit so that it can be visualized by the reviewer
    ii. That informed consent will be obtained (this should be the first thing listed)
    iii. A list (and brief description, if not obvious) of all procedures that will occur during screening
    iv. The time requirement to complete screening

    For retrospective studies, subject screening is performed by evaluating inclusion/exclusion criteria in the existing data source, and can be covered in a sentence or two. A waiver of informed consent is usually requested in retrospective studies (more on this in Session 4).

C. ***Randomization Visit (or Intervention Visit).*** This is the visit where the first of any intervention to be delivered in your study occurs. It should be labeled separately, even if it occurs on the same day (or essentially right after) the Screening Visit. It is important that you include the following in this section:
    i. Any narrative needed to provide context or flow of the visit so that it can be visualized by the reviewer
    ii. A verbal description of the intervention and/or randomization groups, including the control intervention. If some sort of medical procedure is being done as the research intervention, there should be some description about how that procedure is performed (again, enough detail so that risks of the procedure may be judged).

[1] Hulley, SB et al. 2007. Designing Clinical Research, 3rd edition. Philadelphia, PA: Lippincott, Williams, & Wilkins.

iii. If randomization will occur, a brief mention of how randomization will be done. A random number table? The random number function in Excel? Remember that drawing from a fixed number of pre-labeled assignments from a bag is not true randomization, since the odds of receiving a given intervention changes with each assignment; this approach is called block randomization and may be appropriate in certain circumstances.

Observational studies and retrospective studies will not have this type of visit.

D. ***Visit 1, Visit 2, etc.*** The remainder of visit descriptions will depend on your study design. These visits might include additional intervention delivery, data measurement, or both. From reading all of the study visits, it should be clear how you will collect all of the data you need to assess your outcomes. Visits and procedures should be justified by your outcomes; for example, if your outcome is radiographic appearance at 6 months, there should not be both 6-month and 1-year visits for a radiographs (you would need to change your outcomes to justify both measures). It is important that you include the following in this section:
   i. Any narrative needed to provide context or flow of the visit so that it can be visualized by the reviewer
   ii. A list (and brief description, if not obvious) of all procedures that will occur during the visit
   iii. The time requirement to complete each visit
   iv. Your final visit should describe any plans for subject continuity that might be required. For example, if a study treatment will be discontinued, that subjects will be referred to their PCP to ensure their condition remains treated. Another example is if the outcome of your study relates to some sort of new diagnosis, ensuring the subject has education and a referral plan for this new possible diagnosis.

Retrospective studies will not have this type of visit.

E. ***Follow-up Visit.*** A follow-up visit is used as an additional data collection period following an appropriate period after the main outcome collection study visits are completed. This type of visit is more typical for a study that has had an ongoing intervention (like a drug treatment); the follow-up visit typically occurs after the study intervention is discontinued. Therefore, this visit is not as much about efficacy outcomes, but more about safety outcomes following withdrawal of therapy or final study procedures. You can consider a follow-up visit in any study where you might have a lingering concern about your subjects' well-being; e.g., interventions with potential delayed adverse effects, or an invasive procedure, at the final study visit.

Observational studies and retrospective studies will not have this type of visit.

Retrospective Studies: The narrative of the procedures section should discuss how you will access and record the data, and what data variables you will record. A full listing of the variables being recorded is usually needed, and it should be clear why each one is important to achieve the study aims. A diagram/table is rarely useful for

[1] Hulley, SB et al. 2007. Designing Clinical Research, 3rd edition. Philadelphia, PA: Lippincott, Williams, & Wilkins.

retrospective studies. Since the risk of a retrospective study is informational risk, the procedures need to be descriptive enough for a reviewer to gauge the informational risks – i.e., what data are collected, why, and how they are stored.

[1] Hulley, SB et al. 2007. Designing Clinical Research, 3rd edition. Philadelphia, PA: Lippincott, Williams, & Wilkins.