Understanding Clinical Trial Design: A Tutorial for Research Advocates
Understanding Clinical Trial Design: A Tutorial for Research Advocates

Authored by Jane Perlmutter, PhD
for Research Advocacy Network’s Advocate Institute

Table of Contents

I. Introduction ..............................................1
   About the Tutorial .....................................1
   Trade-offs in Designing Clinical Trials ..............2
II. Questions to Ask about Clinical Trials ...............3
   What Research Questions Are Being Addressed
   In The Trial, And How Important Are They? ........4
   Why Should The Trial Be Conducted—i.e., Does The
   Scientific Rationale Adequately Support the Research
   Questions? .............................................5
   Who Will Be Involved In The Trial And How Likely
   Are They To Embrace It? ..............................5
   How Well Designed Is The Trial To Answer The
   Questions It Addresses? ...............................6
   PICO: Patient Issues ..................................6
   PICO: Intervention Issues .............................7
   PICO: Comparison Issues .............................8
   PICO: Outcome Issues .................................8
III. Introduction to Clinical Trial Design .............11
   The Scientific Method Applied to Medicine ........11
   Randomized Controlled Trials .....................14
   Hypothesis Testing and Statistical Inference ....16
IV. Innovations in Trial Design ...........................23
   Introduction to Bayesian Concepts .................23
   Introduction to Adaptive Designs ..................27
      Patient Allocation Adaptive Design ..............29
      Patient Preference Design .......................31
      Patient Enrichment Strategy .....................32
      Randomized Discontinuation Design ............32
      Summary: Adaptive Design .......................34
V. Conclusions ............................................35
VI. Acknowledgements ...................................36
VII. Recommended Reading ...............................37
   Introductory Statistics ..............................37
   History of Statistics ................................37
   Drug Development ..................................37
   Bayesian Statistics ..................................37
   Adaptive Designs ....................................37
VIII. Glossary .............................................38
   Glossary Sources ....................................52

List of Figures

1  Design of Clinical Trials: Striking a Balance ........2
2  Thinking About Clinical Trials ........................4
3  Francis Bacon .........................................11
4  The Scientific Method ................................12
5  Randomized Clinical Trials: The “Gold Standard” ...14
6  Ronald Fisher ..........................................16
7  Introduction of Randomized Controlled Trials ....17
8  Statistical Inference in Hypothesis Testing .........18
9  Errors in Hypothesis Testing .........................19
10 Thomas Bayes .........................................23
11 Paradigm Shift ........................................23
12 Baye’s Theorem .......................................24
13 Frequentist versus Bayesian Approaches ..........26
14 Patient Allocation Adaptive Design ..............30
15 Patient Preference Design ............................31
16 Randomized Discontinuation Design ............33

List of Examples

Concept Development Example ..........................12
Experimental Design (PICO) Example ..................13
Confounding Example ....................................14
Blinding Example ........................................15
Stratification Example ....................................16
Null Hypothesis Example .................................17
Statistical Inference Example ..........................18
Judicial Inference Example ..............................21
Bayesian Sports Example .................................25
Bayesian Diagnostic Example ..........................25
Bayesian Clinical Trial Example .......................26
Adaptation Method Example ............................28
Adaptation Rule Examples ...............................28
Patient Allocation Adaptive Design Example .......29
Patient Preference Design Example ....................31
Patient Enrichment Example ...........................32
Randomized Discontinuation Design Example ......33
I. Introduction

The purpose of this tutorial is to provide a strategy that research advocates can use to constructively contribute to planning clinical trials. It should also assist them to critically assess already designed trials they may be asked to critique (e.g., in grant proposals), as well as to evaluate completed trials (e.g., in journal articles).

The presentation is based on three assumptions about the role of research advocates. First, research advocates have a unique and important contribution to make to clinical research. This is because their focus is primarily on treating patients, rather than on advancing science or careers. Also, their perspective is holistic rather than disease focused. Further research advocates’ energy, sense of urgency and varied experiences outside of research add much value. Second, the most constructive approach research advocates can use to impact research is to raise questions. Raising questions is both less threatening to scientists and less daunting to advocates than providing suggestions. Third, effective research advocates need not be experts in experimental design, statistics, or science. Nevertheless, the more familiar they become with these areas, and the more comfortable they become with the language and style of scientific discourse, the more effective they will be in influencing the course of research.

This tutorial follows from these assumptions. Generic questions that advocates can ask about virtually any clinical trial are presented first. Limited knowledge of clinical trials is required to tackle this section, but by its conclusion readers who do not already have a conceptual framework for thinking about clinical trials should have obtained one (c.f., Figure 2). Additionally, a glossary that contains italicized terms is provided to support readers of varying backgrounds.

For many readers, the section devoted to questions will be sufficient. However, because advocates’ input is enhanced by their understanding of the underlying science, additional background on clinical trial design will also be provided. First the underlying logic of traditional clinical trials is described. This discussion focuses on randomized controlled trials which are the basis of approval of most new medical treatments. It will include an introduction to hypothesis testing and basic statistical concepts. While not essential to research advocates, familiarity with these concepts will help them understand and engage in discussions of clinical trials. Next, a brief introduction to innovative approaches to clinical trial design will be presented. This will include discussion of Bayesian approaches and adaptive designs.

Trade-offs in Designing Clinical Trials

Research advocates are increasingly playing an important role in designing clinical trials that are patient focused and likely to lead to important changes in clinical practice. We want to be sure that clinical trials are designed in a way that will lead to unequivocal results (i.e., are effective at answering research questions). However, we also want to be sure that trials can be completed as rapidly and inexpensively as possible (i.e., efficiently use resources), and that the patients who volunteer to be in trials get the best possible treatment (i.e., the trials achieve the highest ethical standards). These goals are often at cross purposes; thus, clinical trial designs generally represent a compromise. As shown in Figure 1 (page 2), in
addition to these trade-offs, trial designs must balance the priorities of many stakeholders, including trial sponsors, funders, regulators, principle investigators, research collaborators, and community healthcare providers.

**Figure 1. Design of Clinical Trials: Striking a Balance**

Researchers are most concerned with the soundness of the science. They are trained to be methodical, and because their work builds upon previous trials, they place a premium on maintaining a strong scientific foundation. To protect against making errors, their trial designs typically require a large number of patients and a long time to reach conclusions. They also cost a great deal of money. Even so, for a variety of reasons many trials never lead to definitive conclusions. For example, trials often close prematurely because of slow accrual or loss of funding. Also, unanticipated problems with the selection of patients, procedures used during the trial, or very small differences between the interventions being compared can result in inconclusive trials.

Like researchers, informed research advocates should place a high premium on sound science. However, we also need to keep the urgency to rapidly find new treatments front and center. There is good reason to believe that innovative approaches to designing clinical trials can often reduce the time and money needed to successfully complete them. Such approaches can also improve the treatment patients receive on trials, while at the same time maintaining the highest standards of sound science. By asking the right questions, research advocates can encourage researchers to be more innovative in their trial designs.
II. Questions to Ask About Clinical Trials

This section should assist research advocates to formulate useful questions to raise about clinical trial design. The questions are generic, but not exhaustive. Asking generic questions is often possible and even preferable to raising specific suggestions, especially for research advocates whose primary role is to maintain focus on all aspects of patients’ lives. Most of the questions require limited background in clinical trial design, whereas the underpinnings for the more challenging questions are provided in the remainder of this tutorial.

Perhaps the most important questions to ask are:

Will the study design thoroughly address the primary question of the trial?

What alternative trials designs were considered and why was this one selected?

Having researchers articulate the answer to this question has a number of benefits.
- It focuses attention on the primary question the study is supposed to answer.
- It ensures that researchers have seriously thought about the strengths and weaknesses of their chosen trial design relative to alternatives, and requires them to clarify and weigh the alternatives.
- If the question is asked in the presence of other researchers, it opens up discussion among knowledgeable people who may have different opinions on the topic.
- It helps research advocates better understand trial design.
- It helps the research advocate explain and justify the trial to potential participants and supporters of the trials.
- It gives researchers practice at discussing clinical trials in ways that will be understandable to patients they will recruit for their trials.

In the remainder of this section four generic questions will be developed which will help organize thinking about clinical trials. As shown in Figure 2, they relate to the what, why, how, and who of clinical trials. The questions are:

1) What research questions are being addressed in the trial, and how important are they?

2) Why should the trial be conducted—i.e., does the scientific rationale adequately support the research questions?

3) Who will support the trial and how likely are they to embrace it?

4) How well designed is the trial to answer the questions it addresses?

Figure 2 (page 4) also indicates the key components underlying each of these questions, as well as the section of a grant proposal, trial protocol or journal article in which information relevant to each question can typically be found.
Figure 2. Thinking about Clinical Trials

<table>
<thead>
<tr>
<th>Topic</th>
<th>Over-Arching Question</th>
<th>Key Components</th>
<th>Where Addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>What?</td>
<td>What research questions are being addressed in the trial, and how important are they?</td>
<td>• Historical Context&lt;br&gt;• Clinical Importance</td>
<td>Aims</td>
</tr>
<tr>
<td>Why?</td>
<td>Why should the trial be conducted — i.e., does the scientific rationale adequately support the research question?</td>
<td>• Pre-trial Data&lt;br&gt;• Biologic Processes</td>
<td>Background</td>
</tr>
<tr>
<td>Who?</td>
<td>Who will support the trial and how likely are they to embrace it?</td>
<td>• Physician’s Perspective&lt;br&gt;• Clinical Sites&lt;br&gt;• Patient Accrual&lt;br&gt;• Patient Retention</td>
<td>Appendix</td>
</tr>
<tr>
<td>How?</td>
<td>How well designed is the trial to answer the research questions it addresses?</td>
<td>• Patients&lt;br&gt;• Intervention&lt;br&gt;• Comparison or Control&lt;br&gt;• Outcome</td>
<td>Method</td>
</tr>
</tbody>
</table>

Raising some of the more specific questions presented in the remainder of this section will allow research advocates to engage in discussions with researchers to ensure that trial designs have adequately addressed issues that are important to patients.

**What Research Questions Are Being Addressed In The Trial, And How Important Are They?**

Research advocates are constantly aware of the limited number of patient volunteers and other resources that are available for clinical trials, as well as the urgency to make rapid progress in discovering better treatments. Thus, in assessing any trial, we should try to determine its potential to change clinical practice, compared to alternative trials that might be conducted.

In the best case scenario, what information will be gained from this trial?
• How likely is this trial to lead to changes in clinical practice?
• Will this knowledge still be relevant by the time the trial is likely to be completed?
• How many future patients are likely to be impacted by the results of this trial?

What are the opportunity costs associated with this trial?
• What alternative trials are competing for patients and other resources that will be involved in this trial?
• Are there alternative, less costly or less time consuming ways to obtain the same knowledge?

How useful will this trial be to future researchers?
• Will patients be followed after the trial with the goal of collecting evidence about secondary endpoints, and long-term side effects (e.g., secondary cancers, cognitive deficits, survivorship issues)?
• Will patients’ bio-specimens (e.g., blood, tumor tissue) be banked so that it can be used to shed light on biomarkers that may be of interest to future researchers?
• Are the processes by which bio-specimens are collected, handled and stored adequately specified to ensure that they will be useful to future researchers?
Why Should The Trial Be Conducted—i.e., Does The Scientific Rationale Adequately Support The Research Questions?

A key component of success of a clinical trial is the strength of the underlying science. Thus, it is important to raise questions about both pre-trial data and underlying biological processes. Researchers should be willing and able to answer these questions in ways that are understandable to research advocates, not to mention the public that often funds their work and the patients who participate in their trials. Although research advocates will not always be in a strong position to evaluate all aspects of the answers they receive to these questions, they will generally be able to differentiate between potential trials that are scientifically well-grounded, versus those with limited scientific basis or muddled logic.

How strong are pre-trial data?
• Has this intervention already been proven in other patient populations (e.g., for cancers in other organ sites or stages)? What makes the researchers believe it will also be effective in this trial?
• Is there strong evidence that the intervention works in an appropriate animal model? What are the strengths and weaknesses of the animal model?
• Is there strong evidence that the intervention works in an appropriate in vitro model? What are the strengths and weaknesses of the in vitro model?

How strong is the underlying biology?
• Is the experimental intervention targeted at a well established biological mechanism?
• How strong is the evidence that this mechanism is important to the disease process?
• How strong is the evidence that the experimental intervention will be effective in modifying this mechanism?

Who Will Be Involved In The Trial And How Likely Are They To Embrace It?

Even if there is good scientific reason to believe that a trial will be scientifically successful, unless patients can be recruited and retained, the trial will not succeed. Indeed, many trials are terminated early because they cannot recruit enough patients. In such cases all resources that went into planning and partially completing the trial are essentially wasted. Thus, it is important to assess the likelihood that the trial can be successfully completed.

How attractive is the trial protocol from the point of view of physicians who are likely to recruit patients?
• Will it be easy to provide the intervention?
• Will it be easy to collect the required data?
• Is there adequate compensation?
• Are other effective interventions available for eligible patients?
• Are other interesting trials available for eligible patients?

What sites are likely to open this trial?
• Who is sponsoring this trial and how strongly will they “market” it?
• Do sufficient patients meet the eligibility requirements to make opening the trial worthwhile?
• Will the trial be available in community settings, or only at research hospitals?
How effective are the accrual and retention plans?
• Are research advocate organizations involved?
• Are there adequate plans to reach out to underserved populations?
• Will excellent patient support materials that are culturally sensitive be provided (e.g., brochures, schedules, videos)?
• Does the informed consent process comply with best practices?
• Is ongoing patient support planned (e.g., scheduling, psycho-social consultation, pain management, peer support)?

How attractive is the trial protocol from a patient’s point of view?
• How effective is the standard intervention? (Note: Patients are less likely to volunteer for clinical trials when effective standard interventions exist.)
• Are alternative clinical trials available? How attractive are they?
• Will the trial extend the length of treatment?
• How many additional hospital stays, doctor visits, procedures, etc. will be required? How intrusive or inconvenient will these be?
• What are the financial consequences of participating in this trial? Which interventions and tests will be covered by the investigator? Will the patients’ insurance companies pay for costs not covered by the investigator? Who will pay for treatment of side-effects? Will patients be compensated for travel and/or other expenses?
• Are physical accommodations available for patients who travel from out of town, or who have long lapses between procedures?
• How strong is the evidence that the experimental intervention will be effective? How much benefit is it likely to have?
• What are the likely and less likely side-effects of the experimental intervention? Are effective treatments available for these side effects? Are the side effects likely to resolve when treatment ends?
• How likely are unexpected long-term side-effects (e.g., secondary cancers, cognitive deficits) from the experimental intervention?

How Well Designed Is The Trial to Answer The Questions It Addresses?

The acronym PICO is used by many researchers to organize the key elements of clinical trial design, and it will be used here. In particular, concepts are discussed and questions raised about each of the four PICO letters—Patients; Interventions; Comparisons; and Outcomes.

PICO: Patient Issues
Questions about which patients will participate in the trial (i.e., eligibility requirements) help establish that the results of the trial will be applicable to the population of patients whose treatment is likely to change if the trial is successful. They also help highlight the value of patients who volunteer to participate in clinical trials, and that like funding, patients are a limited resource.

Are the eligibility requirements optimal?
• What are the pros and cons of making the eligibility requirements (e.g., disease site or stage, biomarkers, co-morbidities, prior treatments) more or less stringent?
• Would changes to the eligibility requirements increase participation of patients from underserved populations?
• How well do the eligibility requirements match the likely clinical use of the intervention, if the trial is successful?
• Is there an adequate patient population from which to accrue to this trial?
Does this design make the most efficient use of patients?
• Could the primary questions be answered with fewer patients?
• Is there a way to design the trial so that fewer patients are exposed to the less effective interventions (cf., section below on patient allocation adaptive design)?
• Is there a way to design the trial so that results could be achieved more rapidly?
• Could additional secondary questions be addressed without compromising the study?
• Can the study be modified to determine not only whether the intervention is beneficial, but also which patients are most likely to benefit?

How will patients be treated upon completion of the trial?
• Will patients who participated in the trial but did not receive the experimental treatment have subsequent access to the experimental treatment if it is found to be effective?
• Will the experimental intervention be made available to patients who did not receive it, if it is found to be effective?
• What long-term follow-up is planned?
• Will patients be informed about the results of the trial?

PICO: Intervention Issues
The goal of a clinical trial is to determine the impact of an experimental intervention (used interchangeably with investigational intervention). When researchers plan the intervention, they focus primarily on its potential impact on the disease. When research advocates think about the intervention, on the other hand, we consider its impact on all aspects of patients’ lives. This is important because patients who volunteer to participate in clinical trials deserve not only to receive excellent care, but also to be minimally inconvenienced. Further, from a practical point of view, aspects of the intervention that may have limited relevance to the disease (e.g., number of clinic visits), are important to patients and may impact trial accrual and retention, both of which are crucial for the success of trials.

Why was the experimental intervention selected?
• If this is a drug trial, what drug, dose, and schedule of administration will be used? What alternatives were considered?
• If this is not a drug trial (e.g., radiation, surgery, psycho-social, quality of life, correlative science intervention) what variations on the interventions were considered?
• What supportive therapies (i.e., drugs provided to counteract side effects) will be provided and under what circumstances?
• Under what circumstances will the drug dose or other aspects of the intervention be modified?

Are all of the test procedures (e.g., blood draws, scans, biopsies) necessary?
• Are less intrusive procedures available?
• How time consuming will these procedures be?
• Must all of the procedures be completed at the research center?
• Can the procedures be scheduled in a way that minimizes the number of trips a patient must make to the research center?
PICO: Comparison Issues
Assessing an experimental intervention requires comparing it to a comparison intervention (used interchangeably with control intervention). Questions about the nature of the comparison help establish that the trial is ethical. For example, for serious diseases for which useful therapies exist (e.g., many cancers), it is unethical to use placebo comparisons; rather comparison groups (used interchangeably with arm) typically receive the standard of care.

Additionally, to be able to conclude that the experimental intervention differs from the control requires that patients receiving different interventions are otherwise equivalent. Put another way, it is important to avoid any bias or confounding that might provide alternative explanations of intervention effects. Researchers generally focus on eliminating sources of bias that are related to the disease (e.g., stage of disease, prior treatment), whereas research advocates who think more holistically about patients often identify sources of bias that researchers may overlook (e.g., likelihood of remaining in the trial or complying with the protocol).

Is the control intervention appropriate?
- Is there a standard of care that will be prescribed, or will physicians be allowed to choose among interventions?
- Will researchers, health care providers, or patients know to which intervention arm patients were assigned (i.e., Is there blinding)?
- Will tests be performed on patients in both the experimental and control arms, even if they are not part of standard care? (Note: This provision is typically necessary to ensure blinding)

How will patients be allocated among intervention arms?
- Are there ways in which patients assigned to different interventions arms may systematically differ (e.g., demographics, stage of disease)?
- What, if any, patient attributes (e.g., gender, disease site or stage) will be stratified? How were these factors chosen?
- What demographic and baseline variables will be measured to ensure that all groups were indeed equivalent?

How will data be analyzed when the standard protocol is not followed?
- If patients do poorly in the group to which they were assigned, will they be allowed to crossover?
- How will the statistical analysis deal with patients who crossover or drop-out of the trial? (i.e., question whether analysis is “intent-to-treat” or “what was actually received.”)

PICO: Outcome Issues
Clinical trials assess the effect of different interventions on the course of disease by measuring specific outcomes. The choices of outcomes or endpoints typically involve trade-offs that reflect priorities concerning speed, completeness, and clinical value. Primary endpoints (e.g., overall survival, disease free survival, proportion of responders) that are of highest interest are selected and the trial is designed to ensure that they can be adequately assessed. Additionally, secondary endpoints of lesser interest are specified in the protocol (e.g., side-effect profile, quality of life—QOL), but the trial may not be powered to adequately assess them.
What is the primary endpoint?
- Is it important to patients?
- Will it allow the trial to rapidly lead to results?
- Will it lead to definitive results?
- How will it be measured?
- Is the measure reliable and valid?
- What alternatives were considered and why was this one chosen?

Surrogate endpoints are outcomes that have been shown to be early indicators of clinical outcomes that are of interest (e.g. overall survival). An example is cholesterol which has been shown to predict heart attacks, and to be mechanistically related to them (i.e., by blocking arteries). The advantage of using surrogate endpoints is that they are available sooner than the outcomes for which they are surrogates, and hence allow trials to complete more rapidly and less expensively. Adequately demonstrating the appropriateness of a surrogate is, however, difficult. In trials that propose to use surrogate endpoints research advocates should ask:

Why was this surrogate endpoint selected?
- What clinically relevant outcomes are correlated with the surrogate?
- What is the evidence that impacting the surrogate endpoint will also impact the clinical outcome that is of primary interest?
- How will the surrogate endpoint be measured?
- Is the measure reliable and valid?

Additionally, many current clinical trials include the collection of a host of demographic and biomarker measures (sometimes referred to as secondary endpoints) that are analyzed in the hope of identifying questions worthy of future study. Analysis of these variables is called correlative science. Making sure that the correlative science associated with a clinical trial is as effective as possible could have large effect on future progress, and is worth probing in detail. While detailed consideration of these issues is beyond the scope of this tutorial, several basic questions to ask about the correlative science follow.

What other variables (e.g., biomarkers, side effects, cognitive status, quality of life (QOL) and demographic variables) will be of measured?
- How were they chosen?
- What alternatives were considered and why were these chosen?
- If biomarkers will be measured from bi-specimens (e.g., blood, tumor tissue), how will the bio-specimens be collected and handled? How will the biomarkers be assayed?
- How will they be measured?
- Are these measures reliable, valid, and clinically important?
- How will these data be analyzed and used?

Finally, a question always worth asking experts, not just about clinical trial designs, is:

What other questions should I ask?
While not essential, many research advocates are motivated to achieve a deep understanding of the research they try to influence. The purpose of this section is to present the logic and foundation that underlies clinical research. First the scientific method and its application to medicine are described. Then the key components of randomized controlled trials are discussed. The final subsection introduces hypothesis testing and some basic statistical concepts that are used by researchers to provide confidence in the inferences they draw from clinical trials.

The Scientific Method Applied To Medicine

Evidence-based medicine depends on the systematic accumulation of information about how different treatments affect patients. Ideally, a cause-effect relationship can be established between treatments and outcomes in patients with specific diseases. Francis Bacon (Figure 3) is often credited with being the father of the modern scientific method, which is the system underlying evidence-based medicine. It is based on inductive methods that can be used to draw general conclusions based on limited observation, in other words, using observations from a patient sample to draw conclusions about its patient population.

The scientific method is schematized in Figure 4 (page 12). The four key iterative stages are shown in the center, blue boxes:
1) Observe Stage which can entail both formal and informal observation.
2) Hypothesize Stage which articulates the research question in a testable format.
3) Test Stage which entails experimentation. Clinical trials are experiments that involve patients.
4) Conclude Stage that by validates or modifies the hypothesis. The conclusion generally leads to additional observation and experimentation.

The green clouds on the right side of Figure 4 provide examples of activities involved in each stage. Informed by their unique patient experiences, research advocates participate in all of these activities. The pink clouds on the left side of Figure 4 are the processes involved in moving through the four stages of the scientific method.

1) Concept Development Process assimilates observations from a variety of sources and frames a formal research question and testable hypothesis. In clinical research, this process often results in a trial concept document. The “what” and “why” questions raised in the previous section are particularly relevant to this process.
2) Experimental Design Process translates the research question about a population of interest into a formal experiment or clinical trial protocol. The protocol includes patient eligibility requirements, detailed descriptions of the experimental and control interventions, as well as definition of objective, measurable outcomes. The PICO questions raised in the preceding section are especially relevant to this process.
3) Statistical Inference Process is the process that allows researchers to draw conclusions about their hypotheses. The subsections below on “Hypothesis Testing and Statistical Inference” and on “Introduction to Bayesian Concepts” provide two alternative approaches to statistical inference.
Decisions about whether or not to proceed with the research are made between each stage in this schema. For example, research advocates have come to play an especially important role in reviewing grants that allocate funding; this occurs between the “hypothesize” and “test” stages.

Throughout the remainder of this tutorial a hypothetical clinical trial for patients with cancer of the big toe will be used. This example is used because it is concrete, but avoids the distractions associated with more realistic trials.

Consider a clinician who treats patients with cancer of the big toe. In talking to a subset of her patients who had especially favorable outcomes (i.e., survive longer than typical), she noticed that many of them were big coffee drinkers. This raised the research question: “Does coffee drinking improve the survival of patients with cancer of the big toe?”

Research questions typically involve the four PICO components with the patient component is stated in terms of a target population that includes all current and future patients. Also, the comparison is often implicit.

The research question is translated into a testable hypothesis such as: Patients with cancer of the big toe, who drink coffee, survive longer than those who don’t.”
Experimental Design (PICO) Example

An experimental design that can be used to test the research hypothesis is articulated in a clinical trial protocol. The protocol specifies the eligibility requirements of the sample of patients who will be studied, which is presumed to be representative of the population of interest in the research question. The protocol also provides details about the other PICO components. The details can be used by other researchers to interpret the results of clinical trials, or to replicate them.

For the cancer of the big toe example, the following design characteristics will be used throughout this tutorial.

| Patients                  | • Sixty patients (thirty each in the experimental and control arms)  
                           | • Stage III cancer of the big toe                                   |
|---------------------------|---------------------------------------------------------------------|
| Intervention             | • 500 mg. of caffeine                                               
                           | • Administered orally                                              
                           | • Twice a day for 90 days                                          |
| Comparison               | • 500 mg. of colored water                                          
                           | • Administered orally                                              
                           | • Twice a day for 90 days                                          |
| Outcome                  | • Overall survival                                                  |

In many clinical trials the control arm receives no intervention or a placebo. However, because of the seriousness of the disease, in cancer clinical trials patients in the control arm typically receive the current standard of care, if there is one. Patients in the experimental arm, on the other hand, typically receive the same intervention as the control arm, plus an additional experimental intervention, or an experimental intervention that is expected to be at least as effective as the standard of care.

In practice, clinical trial protocols go through many reviews and revisions (e.g., Institutional Review Boards) prior to opening for patient accrual, often including input from research advocates. In the conduct of a trial, circumstances often prevent perfect adherence to the protocol. For example, patients may skip a day of treatments, drop out of the trial, or some of their data may be missing. However, if the trial was carefully designed and run, statistical inference allows researchers to draw conclusions about the research question. This inferential process will be described below. First, however, a brief review of randomized clinical trials, the most common and useful trial designs will be presented.
Randomized Controlled Trials

Randomized Control Trials, described in Figure 5, have become the gold standard of clinical research. To establish causality between the intervention (i.e., caffeine) and the outcome (i.e., overall survival), researchers assume and take steps to ensure that the experimental and control arms are similar in every way except the interventions. This is sometimes referred to as balancing the groups, and ensuring that no superfluous variables are confounded with the intervention. Three techniques to avoid confounding will be discussed.

**Figure 5. Randomized Clinical Trials: The “Gold Standard”**

<table>
<thead>
<tr>
<th>Experimental/Investigational Group Arm</th>
<th>Comparison/Control Group Arm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental or Investigation Treatment, plus Standard of Care</td>
<td>Standard of Care</td>
</tr>
</tbody>
</table>

- Equal number of patients randomly assigned to two or more treatment arms
- Triple blinded (patients, healthcare providers, and researchers), if possible
- Single primary endpoint
- May require many trials to answer complicated questions

1) **Randomization** assigns patients to treatment arms by chance, avoiding any systematic imbalance in characteristics between patients who will receive the experimental versus the control intervention. Usually patients are assigned equally to all arms, although this need not be the case. With a simple two-arm trial (one experimental and one control) randomization can be accomplished with a flip of a coin. When there are more than two arms, or unequal numbers of patients are to be assigned to different arms, computer algorithms can be used to ensure random assignment. The following example demonstrates the importance of randomization.

**Confounding Example**

Consider a clinical trial in which overall survival is the outcome of interest. Suppose a large proportion of patients assigned to the experimental intervention have earlier-stage disease than patients assigned to the control arm. In this situation disease stage and intervention are said to be confounded.

Now suppose that patients who received the experimental intervention lived longer than patients who received the control intervention. Is the survival difference because the experimental intervention is truly better than the control? Or is it because patients in the experimental arm were healthier to begin with? There is no way to determine which explanation is correct.

The difference in prognosis between arms in this trial could have arisen from many subtle biases in how patients were assigned to the experimental versus control arms. For example, healthcare providers may unconsciously assign sicker patients to the control intervention because they have more experience dealing with its side-effects. If patients were randomized, however, imbalances in disease stage would be highly unlikely, especially in large trials.
2) **Blinding** is ensuring that neither patients, healthcare providers, nor researchers know to which group specific patients are assigned. Trials are said to be single, double, or triple blinded, depending upon how many of the relevant participants in the trial are unaware of patient assignment. The purpose of blinding is to minimize patients receiving different care, or their data be interpreted differently, based upon the intervention they are assigned. The following example demonstrates the importance of blinding.

**Blinding Example**

Consider a clinical trial in which it is suspected that an experimental intervention delays relapses compared to the control intervention, but is also more toxic. If this trial was not blinded, patients in the experimental arm might be especially vigilant to report toxicities. Likewise, their healthcare providers might unwittingly monitor these patients more closely than if they were assigned the control intervention.

Now suppose the experimental intervention was found to be more toxic than the control. Would this be due to a real difference in toxicity or reporting difference? There is no way to know. Had the patients and healthcare providers been blinded to which arm patients were assigned, this ambiguity would not have arisen?

3) **Stratification** prior to randomization can be used to ensure that the number of patients assigned to the experimental and control arms are balanced with respect to important attributes (stratification variables). Examples of stratification variables are gender or disease stage. The purposes of stratification are two-fold. First, stratification ensures that the stratification variable is not confounded with the intervention, which is especially important when the stratification variable is known to have a large impact on the outcome of interest. In large trials randomization alone typically achieves balance, and stratification may be unnecessary. Second, if adequately planned, stratification allows sub-group analysis, essentially looking at each stratum separately. It is not uncommon, however, for subgroup comparisons to be conducted even when not adequately planned. This leads to increased error rate, as discussed in the section on "Hypothesis Testing and Statistical Inference" below. The following example shows how stratification can help ensure that the results of a clinical trial can be interpreted.
Stratification Example

Consider a clinical trial that is designed to test a new treatment believed to be beneficial for all solid tumors. The trial might include patients with lung, colon, and breast cancers. However since the disease process is so different among these tumors, patients might be stratified by disease site. In particular, patients with lung cancer will be randomized to either experimental or control arms, patients with colon cancer would be randomized to either experimental or control intervention arms, and patients with breast cancer would be randomized to either experimental or control arms.

In this example disease site is the stratification variable. It was selected because the natural disease course is known to be very different among these tumors. If randomization alone were used to assign patients to interventions, it is possible that a larger proportion of lung cancer patients (shortest average survival) might have been assigned to the control arm. This is fairly likely in small trials, but relatively unlikely in large trials.

In this example, if the control arm which included more patients with poorer prognoses had shorter survival, it would be impossible to determine whether it was due to the difference in treatments or the difference in patients. Prognosis aside, many interventions are found to be effective in only a subset of solid tumors, and stratification may make it simpler to identify such differences.

Detailed discussion of the drug development process is beyond the scope of this tutorial. Nevertheless, the various phases (0-IV) are defined in the glossary, several readings are recommended on this topic, and a few comments are in order here. First, phase III trials that are designed to prove the efficacy of new treatments all use randomized controlled designs, although many employ embellishments of the simple two-arm design discussed here. For example, there may be more than one experimental arm, perhaps with each group using a different new agent, or with each group using the same agent but at different doses. Second, given the different goals of phase I and II trials, alternative designs are often more appropriate. In particular, phase I trials are designed to determine safe drug doses and do not include control groups. Also, phase II trials are designed to establish drug activity. As a result, they are often small and preliminary, and often use historical controls rather than including a randomized control arm.

Hypothesis Testing and Statistical Inference

After a randomized controlled trial is conducted, statisticians help determine whether any observed difference between outcomes in the experimental and control arms are real, or simply chance occurrences. Ronald Fisher (Figure 6) is often credited as the father of modern statistical analysis of experiments, in particular with defining the rules of inference that are used to assess the outcomes of randomized controlled trials. Interestingly, his methods were first applied in the 1920s to problems in agriculture, and are only today impacting educational practice. They began to be widely applied in medical research in the 1960s. This was largely due to the 1962 Kefauver-Harris Drug Amendment which required proof of efficacy, in addition to proof of safety for drug approval. Figure 7 shows the key components of randomized controlled trials in each of these disciplines.
The process to which Fisher’s methods can be applied is called *hypothesis testing*. In *hypothesis testing* a *null hypothesis (H₀)* is articulated which is typically a statement of no difference between *experimental* and *control* patient populations. Articulation of the *null hypothesis* takes place in the first, concept development process of the scientific method described above.

The process to which Fisher’s methods can be applied is called *hypothesis testing*. In *hypothesis testing* a *null hypothesis (H₀)* is articulated which is typically a statement of no difference between *experimental* and *control* patient populations. Articulation of the *null hypothesis* takes place in the first, concept development process of the scientific method described above.

### Null Hypothesis Example

The *null hypothesis* associated with the cancer of the big toe examples is: “There is no difference in the overall survival of patients with cancer of the big toe that are treated with caffeine versus treated with colored water.”

More generically, the *null hypothesis* is typically: “There is no difference in the outcomes of patients treated with the experimental versus control interventions.” The *null hypothesis* is about patient populations—all current and future patients with specific conditions who receive specific interventions. *Statistical inference* is used to decide whether or not the *null hypothesis* is true, based on a *sample* of patients in a *clinical trial*.

Next, the *clinical trial* is designed and run; this is part of the second, *experimental design* process of the scientific method described above. The third process in the *scientific method—statistical inference*—is outlined in Figure 8 (page 18). *Statistical inference* allows researchers to determine whether any observed difference between outcomes in the *experimental* and *control arms* reflects a true difference, or is simply a matter of chance.

<table>
<thead>
<tr>
<th>Widespread Introduction</th>
<th>Agriculture</th>
<th>Medicine</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1920s</td>
<td>1960s</td>
<td>2000s</td>
</tr>
<tr>
<td>Samples</td>
<td>Plots of Land</td>
<td>Patients</td>
<td>Students, Classrooms, Schools</td>
</tr>
<tr>
<td>Example Interventions</td>
<td>Fertilizers</td>
<td>Drugs</td>
<td>Teaching Methods</td>
</tr>
<tr>
<td>Example Outcomes</td>
<td>Crop Yield</td>
<td>Survival</td>
<td>Student Learning</td>
</tr>
</tbody>
</table>
When clinical trials data are analyzed, the outcomes of each arm is described using descriptive statistics such as mean, median and standard deviation. In addition, a test statistic (e.g., t-test or F-test) is computed and compared to the value this statistic would take, if the null hypothesis were true. The final step in hypothesis testing is deciding whether or not to reject the null hypothesis. When the computed test-statistic is different than what would have been expected if the null hypothesis were true, it is rejected; otherwise it is not.

Consider the cancer of the big toe example outlined on page 12. The following table provides descriptive statistics that might have resulted from this trial.

<table>
<thead>
<tr>
<th></th>
<th>Experimental (Caffeine) Arm</th>
<th>Control (Water) Arm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Mean Overall Survival</td>
<td>8.6</td>
<td>7.8</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.6</td>
<td>.5</td>
</tr>
</tbody>
</table>

\[ t = 5.61, \ p < .01 \]

In addition to describing the results, a t-test statistic might be computed from the trial data. In this example, the t value is 5.61. This t-test would be compared to t-tests that result from the null hypothesis in trials with 30 patients in each of two arms. These values can be found in statistical tables. In particular, the tables indicate that a t-test \( \geq 2.65 \) would have occurred by chance \( \leq 1\% \) of the time, if the null hypothesis were true. Since the t-test calculated from the trial data (5.61) is larger, the null hypothesis would be rejected, and researchers claim that survival is greater for patients who are treated with caffeine than with water.
The calculation of test-statistics is beyond the scope of this tutorial. However, three factors always influence their values, and hence how likely they are to deviate from test statistics that would have been computed had the null hypothesis been true. These are:

1) **Sample Size**: The larger the sample, the more likely the observed outcomes reflect their overall populations. The limiting case is when the entire population is part of the clinical trial. Thus, other factors being equal, the larger the sample size, the more likely the null hypothesis will be rejected.

2) **Variability**: The less variability among patients within each group, the more likely they reflect the overall populations. In trials with low variability, trial outcome differences between experimental and control arms are likely to be real (i.e., not due to chance). Thus, other factors being equal, the null hypothesis is more likely to be rejected in trials with low variability.

3) **Outcome Difference**: The larger the difference in outcomes between the experimental and control arms, the more likely there is a true difference, even if it is actually smaller or larger than observed in the trial. Thus, other factors being equal, the larger differences between experimental and control arms, the more likely the null hypothesis will be rejected.

How unlikely would the trial results need to be to reject the null hypothesis? This rests on researchers’ tolerance for errors. The more tolerant of errors, for example, in more exploratory work, the more likely the null hypothesis will be rejected. However, it is never actually known whether or not the null hypothesis is true. Rather, researchers establish criteria to maintain specific error rates across all trials. This is what they do when they set $\alpha$ or type I error rates at .5%, and $\beta$ or type II error rates at 20%. Figure 9 and the following text explain these potential errors in more detail. This material is a rather technical and some readers may choose to skip to the judicial example at the end of this section.

The columns of Figure 9 provide the two possible true states of affairs: either the null hypothesis ($H_0$) is true or false. The rows give the two possible decisions; either fail to reject or reject the null hypothesis. The cells show the conclusions that are drawn—or the experimental and control groups are equivalent or not—and whether the decision was correct (indicated by ☺), or an error (indicated by /).
Decision rules are established to limit the percentage of trials with erroneous decisions. One type of error occurs when the null hypothesis is true but is rejected, (bottom, left cell). This is represented by two ☺☺ in Figure 9 because these errors can lead to serious consequences when new treatments are inappropriately adopted or future research is based on them. These errors are sometimes called false alarms or false positives because they falsely claim interesting differences between the experimental and control groups. They are also referred to as $\alpha$ or type I errors, and decision rules are generally established to ensure they occur in no more than 5% of trials. However, when sub-group analyses are conducted, particularly if they were not planned prior to data collection, $\alpha$ error rates can be much higher.

The other type of error occurs when the null hypothesis is false, but not rejected (top, right cell). These errors are sometimes called misses or false negatives because they occur when interesting, true differences between experimental and control treatments are missed. They are also referred to as $\beta$ or type II errors and decision rules generally ensure that these errors will occur in no more than 20% of trials.

There are also two types of correct decisions. One is when the null hypothesis is false and it is rejected (lower, right-hand cell). In this case, sometimes referred to as a hit, the correct decision is that there is a true difference between the experimental and control arms. This is represented by two ☺☺ in Figure 9 because this is the goal of most clinical trials—to identify new, effective treatments. Clinical trials are designed so that the null hypothesis will be rejected in 80% of cases in which it is false—that is, when $\beta$ errors are not made. This probability of correctly rejecting a false null hypothesis, and in so doing identifying a true treatment difference, is referred to as power, and is always equal to $(1 - \beta)$.

The cells with ☺ reflect the other type of correct decisions—failing to reject the null hypothesis when it is true (top, left-hand cell). In these cases the clinical trial does not provide adequate evidence to conclude that there is any difference between the experimental and control arms. Although many researchers conclude the two treatments are the same, such a conclusion is actually a stronger conclusion than is justified.
It may be helpful to compare statistical inference used in hypothesis testing to the judicial inference process involved in criminal trials. The presumption that the defendant is innocent until proven guilty is equivalent to the null hypothesis. Requiring “evidence beyond a reasonable doubt” to convict a defendant is aimed at minimizing the chances of convicting innocent defendants. This is akin to minimizing type I errors ($\alpha < .05$) in a clinical trial. Further, while there is a desire to avoid acquitting guilty defendants, doing so is somewhat more tolerated. This is akin to the somewhat higher tolerance of making type II errors ($\beta < .20$) by failing to reject a false null hypothesis.

The four possible outcomes of jury trials are compared to the outcomes of clinical trials in the table below. Note that the decision in the clinical trial is always relative to the control (standard of care) intervention.

<table>
<thead>
<tr>
<th></th>
<th>Jury Trial</th>
<th>Clinical Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$</td>
<td>Presumed Innocence</td>
<td>Experimental = Control</td>
</tr>
<tr>
<td>Correctly</td>
<td>Convict a felon intervention</td>
<td>Identify effective</td>
</tr>
<tr>
<td>Reject the $H_0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correctly</td>
<td>Acquit an innocent</td>
<td>Correctly reject ineffective</td>
</tr>
<tr>
<td>Fail to</td>
<td>intervention</td>
<td></td>
</tr>
<tr>
<td>Reject the $H_0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type I Error</td>
<td>Convict an innocent</td>
<td>Claim an ineffective</td>
</tr>
<tr>
<td></td>
<td>intervention is effective</td>
<td></td>
</tr>
<tr>
<td>Type II Error</td>
<td>Acquit felon</td>
<td>Miss an effective intervention</td>
</tr>
</tbody>
</table>

Although the judicial analogy may be helpful in conveying the logic of hypothesis testing, it remains a rather contorted process. An alternative approach based on Bayesian statistics is more natural and gaining influence in medical research. The Bayesian approach will be discussed in the next section of this tutorial.
IV. Innovations in Trial Design

The “Introduction to Clinical Trials” section focused on the current standard practice of clinical trial design. Innovative approaches, while not yet widely used, will be discussed here because they have potential to significantly advance the field of clinical trial design, and lead to more rapid progress in identifying effective cancer treatments. As discussed above, clinical trial design entails balancing competing priorities (Figure 1). Traditional trial designs have maintained the highest priority on avoiding type I errors. To accomplish this, trials are very large and costly, and generally require many years to provide answers. The innovations discussed in this section also place high priority on avoiding incorrect conclusions, but often require fewer patients and dollars to complete the trials, and lead to results more rapidly. Therefore, they should be of interest to research advocates.

Introduction to Bayesian Concepts

Bayesian concepts are not new. Thomas Bayes (Figure 10) lived in the eighteenth century, hundreds of years before Ronald Fisher worked on modern statistical methods. Among other issues (c.f., Winkler, 2001), the computational complexity involved in the Bayesian approach made it impractical to implement his ideas. Given advances in computer technology over the past thirty years, the Bayesian approach has become more practical.

Increasing practicality of using the Bayesian approach, on the one hand, paired with increasing recognition of problems associated with the traditional approach (c.f. Goodman, 2001), on the other hand, are likely to result in a paradigm shift characterized in Figure 11. The different inferential processes that are used and questions that are addressed by traditional (generally referred to as frequentist) and Bayesian approaches are shown in the top half of Figure 11, and will be discussed below. This paradigm shift will significantly impact drug development, as shown in the bottom half of Figure 11. First, drug approval will come to be based on “weight of evidence” rather than “pivotal trials,” as is currently the case. Second, this is likely to lead to a glossing of the phases of drug testing (phase I – III trials). Third, adaptive designs, discussed below, will be the vehicle of this glossing of the phases. Fourth, trials will be analyzed using Bayesian statistics.

<table>
<thead>
<tr>
<th>Inferential Process</th>
<th>Old</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question Being Addressed</td>
<td>Hypothesis Testing (Attempt to reject null hypothesis)</td>
<td>Continuous Learning (Update probabilities of alternative hypotheses)</td>
</tr>
<tr>
<td>Drug Approval</td>
<td>Pivotal Trial</td>
<td>Weight of Evidence</td>
</tr>
<tr>
<td></td>
<td>Distinct Phase 0-IV Trials</td>
<td>Continuous Trials</td>
</tr>
<tr>
<td>Trial Designs</td>
<td>Single Stage</td>
<td>Adaptive</td>
</tr>
<tr>
<td>Statistics</td>
<td>Traditional</td>
<td>Bayesian</td>
</tr>
</tbody>
</table>
The overall conceptual model of the scientific method (Figure 4) holds for both Bayesians and frequentists. Likewise, the issues of randomization and blinding hold for both approaches. However, the Bayesian approach provides an alternative process for carrying out the inferential steps that allow researchers to draw conclusions about populations of patients, based on samples in their trials. The Bayesian inferential process will be described below, and contrasted to the process employed by frequentists that was described in the previous section. Although the frequentist approach has been widely used, the logic is quite contorted. Further, progress has been slow because trials are large and expensive.

**Figure 12. Bayes’ Theorem**

Why are Bayesian trials generally smaller and hence less costly? Bayesians build on prior knowledge, rather than viewing each trial in isolation. Prior knowledge, for example, may be based on trials with similar drugs in different organ sites or disease stages. The concept of incorporating pre-trial knowledge is captured in Bayes’ Theorem which is presented in words in Figure 12. At the start of a trial, a Bayesian will assign a prior probability to the hypothesis of interest, based on the best information available at that time. The trial data will be used to calculate the standardized likelihood, which will be combined with the prior probability to yield a posterior probability, which can in turn be used as the starting point (i.e., prior probabilities) of subsequent trials. In this way, the Bayesian approach is sometimes said to embrace continuous learning.

Three examples of the Bayesian approach are provided below. The first concerns betting on sporting events, and demonstrates that incorporating prior knowledge—in this example, prior success of a football team—is natural to the way people think. The second example is a diagnostic example in which prior knowledge about the prevalence of different diseases, and the specificity of diagnostic tests are taken into account to arrive at the most likely diagnosis. The final example builds on the “cancer of the big toe” example outlined above, and demonstrates how information from prior trials might be incorporated into clinical trials.
Bayesian Sports Example

College football seasons consist of twelve regular season games. At the beginning of the season, a sports fan is generally undecided about the strength of her favorite team. As the season progresses she is likely to become more or less confident of a win, based on the team’s success during previous weeks.

In Bayesian terms, the fan’s prior probability at the beginning of the season was around .5. Following the first game, her posterior probability would be smaller if her team lost the season opener, and larger if they won. That posterior probability computed after game one would be used as the prior probability for considering the second game of the season, and so on throughout the season.

If the fan was a betting person, her money would follow Bayesian logic. For example, if her team won the first ten games, her prior and posterior probabilities would increase throughout the season, and she would be likely to bet on a win in game eleven. On the other hand, if the team had only won half of the first ten games, her prior and posterior probabilities would fluctuate around .5 throughout the season, and she would be hesitant to bet on a win in game eleven.

For illustrative purposes, the success of last year’s season, the strength of each week’s opponents, and whether or not games were played at home, were not incorporated into this example. The Bayesian framework could, however, accommodate them, as well as other relevant information.

The key point is that a team’s football games are not independent events, and thus history influences betting. The same, according to Bayesians, can be said about clinical trials that involve similar diseases or treatments.

Bayesian Diagnostic Example

There are typically many possible causes associated with any set of symptoms. An initial diagnosis is generally the most common or prevalent disease that matches the symptoms. Thus, for example, the common cold is a much more likely diagnosis of a cough than is lung cancer. If the initial diagnosis turns out to be wrong, for example because the cough persisted, tests may be run to refine the diagnosis. In this example, a chest CT might be ordered.

If a suspicious spot is found on the CT scan, a patient may fear the worst and assume she has lung cancer. This concern is especially likely in smokers who have a higher prior probability of having lung cancer.

In actuality, a lung cancer diagnosis would be premature, even in smokers, because the specificity of a chest CT for lung cancer is rather low. That is, there are many causes of spots on lung CTs (e.g., pneumonia) that have higher prior probabilities than lung cancer. Only when these are ruled out would a biopsy, which is a definitive but highly invasive test, be done to establish a correct diagnosis.

To summarize, the prevalence of potential diseases are used to establish prior probabilities. The results of diagnostic tests are factored in to provide posterior probabilities, which can then be used as priors for subsequent tests, until a definitive diagnosis is established.
Bayesian Clinical Trial Example

Consider the hypothetical cancer of the big toe trial outlined on page 14. Researchers hypothesized that the experimental intervention (i.e., caffeine) is superior to the standard of care (i.e., water) in treating cancer of the big toe. In the Bayesian framework a prior probability would be assigned to this hypothesis. For example, based on previous experience (e.g., trials in thumb cancer), they might assign a prior probability of .6 to this hypothesis (i.e., the probability that caffeine is better than water is slightly better than even chance).

After collecting data, the researchers would calculate a posterior probability which becomes their new best guess about the hypothesis. In this example, suppose patients who were treated with caffeine lived an average of 8.6 years following treatment, whereas those treated with water lived an average of 7.8 years. This would provide confirming evidence for the hypothesis, and the posterior probability would be higher than the prior probability, (e.g., perhaps .8). On the other hand, if patients in the control arm survived longer than those on the experimental arm, the trial would cast doubt on the hypothesis. In such cases the posterior probability would be lower than the prior probability (e.g., perhaps .5 or even less).

In this way new beliefs about the hypothesis (i.e., posterior probabilities) combine prior knowledge (i.e., prior probability) with trial data. The larger the trial and the greater the difference between the experimental and control arms, the greater the influence of the trial data on the new belief about the hypothesis.

While the calculations used in the Bayesian approach are beyond the scope of the current tutorial, the above examples should have provided the reader with a sense of how existing knowledge can be incorporated into the inferential process, as well as the naturalness of the approach. In everyday life as well as the clinical setting, very surprising results are almost always discounted. This is consistent with the Bayesian, but not the frequentist approach.

A major criticism of the Bayesian approach is the apparent subjectivity associated with establishing prior probabilities. Who gets to decide which pre-existing data to incorporate and how heavily to weigh them? While not trivial, there are ways around these issues. First, non-informative prior probabilities could be used. A non-informative prior probability would be equivalent to chance, for example, .5 in a two-arm trial. Second, the same decisions about a hypothesis are often reached for a wide range of prior probabilities. Third, in large trials, new data tend to overpower any influence of prior probabilities.

Other differences between frequentists and Bayesians, besides the use of prior information, are summarized in Figure 13. On the one hand, Bayesians use Bayes’ Theorem to address questions and draw conclusions that are framed as probabilities associated with various hypotheses. Frequentists, on the other hand, use statistical methods (i.e., sampling distributions of test statistics) that support hypothesis testing and limit error rates.
These differences have important consequences. First, Bayesians draw conclusions about hypotheses of interest, rather than the null hypothesis. Second, Bayesians can use same data to assess multiple hypotheses. Third, and perhaps most important, Bayesians do not have the same concerns about error rates that plague frequentists. This difference is because Bayesians’ conclusions are posterior probabilities that are neither right nor wrong, but rather estimates about the correctness of hypotheses. Frequentists, on the other hand, reject or fail to reject the null hypothesis, and are thus either right or wrong. Unfortunately, in any given trial they do not know which. Therefore, frequentists go to great lengths to limit the overall number of errors, especially type I errors. This strategy, in turn, leads frequentists to limit the number of times they look at their data. Bayesians, on the other hand, have no difficulty looking at their data as they are being collected, and even using interim results to modify trials. This technique can have major impact on the value and timeliness of clinical trials, as well as on their efficiency. The opportunity to continuously monitor data and modify trials as they are accruing is the basis of adaptive designs which will be discussed in the next subsection.

By discussing Bayesian approaches with researchers, research advocates can play a role in accelerating the paradigm shift described in Figure 11. While it would be a mistake to advocate an immediate adoption of this approach, it is not too soon to encourage the research community to learn about it. As people become more educated in the uses and abuses of Bayesian methods and develop simple software tools to implement them, their strengths and limitations will become more apparent.

**Introduction to Adaptive Designs**

The term “adaptive design” is used in many different ways. Here it is used to describe any multi-stage trial where later stages are based, in part, on what happened in earlier stages. While a Bayesian perspective is not strictly necessary for adaptive designs, as discussed above these designs are natural to Bayesians who are constantly updating probabilities and comfortable continuously looking at their data. Therefore, adapting trial designs as data accumulate is consistent with their paradigm. This is not really the case for frequentists who draw inferences from isolated trials and control error rates. Still, frequentists are increasingly using multi-stage designs because of their appeal. However, this is straining the traditional paradigm and is likely to eventually give way to the paradigm shift described in Figure 11.
By using information as it accumulates, adaptive designs allow researchers to focus their data collection on issues that require the most attention and/or reduce the overall amount of data they collect.

Vision and hearing tests use adaptive methodologies and provide an intuitive sense of the value of adaptive trial designs. In both of these tests initial assessments are similar for all patients, and are designed to rapidly find the general limits of vision or hearing. Later assessments, however, are individualized and designed to fine-tune the necessary vision or hearing correction.

Likewise, adaptive trials are often designed to first establish general characteristics of the trial, and then to focus data collection where it will be most informative. For example, some adaptive trials begin by comparing several drugs or drug doses and through the course of the trial focus on the two that appear to be most beneficial.

There are many modifications that can be made during the course of an adaptive trial. For example, an adaptive sampling rule can determine how many subjects should be included at subsequent stages of a trial. This rule may be based on accumulating information about accrual rate, sample variance, or availability of funding. Alternatively, a stopping rule would establish under what conditions the trial should be terminated, for example, due to observed efficacy, harm, futility, or safety. Several examples of potential adaptation rules are provided below.

**Adaptive Method Example**

**Vision and hearing tests use adaptive methodologies and provide an intuitive sense of the value of adaptive trial designs. In both of these tests initial assessments are similar for all patients, and are designed to rapidly find the general limits of vision or hearing. Later assessments, however, are individualized and designed to fine-tune the necessary vision or hearing correction.**

Likewise, adaptive trials are often designed to first establish general characteristics of the trial, and then to focus data collection where it will be most informative. For example, some adaptive trials begin by comparing several drugs or drug doses and through the course of the trial focus on the two that appear to be most beneficial.

There are many modifications that can be made during the course of an adaptive trial. For example, an adaptive sampling rule can determine how many subjects should be included at subsequent stages of a trial. This rule may be based on accumulating information about accrual rate, sample variance, or availability of funding. Alternatively, a stopping rule would establish under what conditions the trial should be terminated, for example, due to observed efficacy, harm, futility, or safety. Several examples of potential adaptation rules are provided below.

**Adaptation Rule Examples**

**Sampling Rule:** Compute the standard deviation of the sample after the outcome is measured in the first ten patients.
1) If the standard deviation is larger than assumed in planning the trial, recruit more patients than previously planned.
2) If the standard deviation is smaller than assumed in planning the trial, recruit fewer patients than previously planned.

**Stopping Rule:** Compute the test statistic after the outcome is measured in the first ten patients.
1) If the test statistic suggests that the experimental intervention is inferior to the control intervention, with less than a 5% chance of this being due to chance, stop the trial for futility.
2) If the test statistic suggests that the experimental intervention is superior to the control intervention, with less than a 1% chance of this being due to chance, stop the trial due to efficacy.

Whatever the adaptation rules, they must be specified prior to starting the trial. This requires a considerable amount of time and thought should be given to planning adaptive designs, but these upfront costs are typically more than made up by the time the trial is completed.

The remainder of this section will present three specific multi-stage designs that in varying ways, use information obtained in early stages of the trial to modify later stages. They are: 1) patient allocation adaptive design; 2) patient preference design and 3) randomized discontinuation design. In addition, patient enrichment strategies, which can also increase the efficiency of clinical trials, will be introduced.
Patient Allocation Adaptive Design

In the patient allocation adaptive design an adaptation rule modifies the way patients are assigned to intervention arms as a result of accumulating data. Patients are always assigned randomly; what is modified is the proportion of patients assigned to each intervention arm. In particular, more patients are assigned to the intervention that is performing better. If the apparent advantage is real, the intervention arms rapidly diverge and the trial can be concluded. However, if the advantage is due to chance, the intervention arms converge. The total number of patients needed to come to a reliable conclusion is determined as data accumulate, but it is typically fewer than with a traditional one stage design. Additionally, compared to the traditional randomized controlled trial, a larger proportion of patients in the patient allocation adaptive design are treated with the superior intervention. Don Berry, Ph.D. of MD Anderson Cancer Center has used this design effectively in a number of applications, and his simulations demonstrate its efficiency in finding “winners” among a group of potential interventions.

Figure 14 (page 30) presents an example of a two-arm trial in which assignment of patients to intervention is modified through the course of the trial. The adaptation rule specifies that at the start of the trial patients should be allocated equally to the two treatment arms. The adaptation rule also indicates when data should be analyzed, and depending on the results, how the allocation ratio should be changed.

In this example, data were analyzed after outcomes from the first eight patients became available. The response rate (i.e., percentage of patients whose tumors shrink following treatment) was twice as large in one of the arms. The adaptation rule specified that in such cases, the ratio of patient assignment should change from 1:1 to 3:1, favoring the superior arm.

As more data are collected outcomes would continued to be compared at pre-specified points. The adaptation rule would specify under which condition the trial should be:
1) terminated due to clear superiority of one treatment;
2) ended due to clear equivalence of the treatments (i.e., futility of finding difference)
3) continued, with the same patient allocation; or
4) continued with a revised patient allocation.

Note that throughout the trial, all patients were randomly assigned to the experimental and control arms based on a randomization algorithm, even though the proportion of patients assigned to each arm shifted through the course of the trial.
A serious limitation to patient allocation adaptive designs is the need for an outcome measure that occurs relatively quickly before many new patients accrue. An example of where these designs might work well is with an outcome measured at a landmark time, such as, tumor response measured at four months after beginning protocol therapy, engraftment measured six months after transplantation, or biomarkers (used as surrogate endpoints) measured at a specific time after receiving drug. Another possible application is with survival endpoints; however, due to the time-sensitive nature of the design, it might be most appropriate for poor prognosis patient subgroups, like those with pancreatic cancer or those receiving salvage therapy for metastatic disease.

Consider an approach to both cancer research and treatment that might be called a “continuous adaptive trial.” All patients could be treated as part of an adaptive trial that includes all treatments that are likely to be at least as effective as the standard of care. As new treatments reach this criterion, they would be added to the set of treatments included in the trial. As data accumulate, treatments that do not perform well would be removed. Essentially, this would be one large trial that included all treatments currently in phase II, III and IV trials, as well as the standard of care. While a revolutionary idea, this would be a sound evolutionary approach to improving the standard of care. Although conceptually simple, it would, no doubt, be difficult to implement. It is unlikely that many pharmaceutical companies would be willing to participate in such trials. Patient safety would need to be addressed and adjustments to regulatory processes would also have to be considered. Still, it may be worth discussing such an approach. There seems to be the potential for significant improvement in the treatment of patients, as well as more rapid research progress.
**Patient Preference Design**

Another literally “out of the box” adaptive design is sometimes referred to as the *patient preference design* (Figure 15). It is motivated, in part, by distaste for randomization among many potential participants in clinical trials. Using this design, patients can agree to be in trials, and then can select to either be randomized or not. The *adaptation rule* for this trial would specify how many patients to recruit and when to stop the trial based on:

1) proportion of patients who select their own *intervention* versus those who choose to be randomized;
2) overall *outcome* differences between the experimental and control arms; and
3) similarities of *outcomes* in patients who select their own *treatment* versus choose to be randomized.

Figure 15 presents an example of a two-arm patient preference design. Patients agree to participate knowing that they will have the option to select their own treatment, among those included in the trial, or be randomized.

Patients who have no *treatment* preference fall into the top row of Figure 15, which is essentially the traditional two-arm randomized controlled trial design. Patients who want to choose their own treatment would not be allowed to participate in a traditional trial, but are included in trials that use patient preference designs (i.e., the bottom row in Figure 15). Including patients who want to select their own treatment allows the trial to accrue more rapidly.

*Statistical analyses* can be used to determine whether the option to select treatment had any influence on outcomes—that is, whether the pattern of outcomes differs in the two rows of Figure 15. If the same treatment is superior regardless of whether or not patients select it, which is most likely to be the case, the trial can be completed with fewer randomized patients than would have been required in the traditional design.

If the difference in outcomes varies depending on whether or not patients select their own treatment, either in direction or magnitude, that finding itself would be of interest. It might, for example, suggest that patients are more compliant with treatments they actively choose. Such a finding would not be possible to detect in traditional randomized controlled trials.
**Patient Enrichment Strategy**

One of the difficulties of cancer research is that even among patients with the same diagnosis (e.g., stage IV breast cancer), different patients respond to different drugs. The challenge then is to identify patients most likely to respond to each new intervention. This requires framing research questions in terms of the sub-group of patients who are most likely to respond, and establishing eligibility requirements that restrict trials to this target group. This strategy is referred to as patient enrichment. As cancer interventions become more targeted, patients who are most likely to benefit from each new intervention can be selected based on biomarkers—for example, DNA or RNA in their tumor or circulating cells, proteins in their blood, or genetic factors that influence the way they metabolize drugs. The following example indicates the advantage of using a patient enrichment strategy when a relevant biomarker is known.

**Patient Enrichment Example**

Approximately one-third of breast cancer tumors over-express a gene called HER2, and as a result have more HER2 receptors on their cells. These tumors grow more aggressively and result in poorer prognoses. Traztuzumab, a drug commonly referred to as Herceptin, is a targeted therapy that binds to HER2 receptors. Rather than testing its efficacy in all breast cancer patients, initial trials were restricted to patients who over-expressed the HER2 gene.

This was a patient enrichment strategy that allowed researchers to demonstrate the benefit of the treatment in a select subset of patients. The drug was approved for use in patients who over-express HER2. Subsequent trials confirmed that this drug has limited efficacy in women who do not over-express HER2. Had the initial trials not utilized a patient enrichment strategy it is unlikely that the drug would have been shown to be effective. This is because it is ineffective in approximately 85% of breast cancer patients—those approximately 65% who do not over-express HER2, as well as 50% of those who do express HER2, but do not respond to the drug.

**Randomized Discontinuation Design**

It is not always possible to predict which patients are most likely to respond to new treatments. The randomized discontinuation design uses a patient enrichment strategy in a two-stage design, even when there is no way to predict which patients are most likely to benefit from the experimental intervention. It is particularly useful in phase II trials where establishing a drug’s activity is at issue.
Randomized Discontinuation Design Example

Figure 16 presents an example of a trial that uses a randomized discontinuation design. Initially all patients in this trial receive the experimental intervention; this is often attractive to prospective volunteers.

The adaptation rule specifies that patients' tumors are measured six weeks after their treatment begins. Further, their treatment may be modified according to the following rule.
1) Patients for whom the experimental intervention appears to be working—i.e., their tumors shrink—continue on the experimental intervention
2) Patients who do not appear to be benefiting from the experimental intervention are switched to the standard of care
3) Patients who have stable disease are randomized to continue the experimental intervention or switch to the standard intervention. These two randomized groups essentially form the traditional two-arm randomized controlled trial.

Evidence of treatment activity can come from two sources in this design. First, if a significant proportion of patients fall into first group above, the treatment is likely to have value. The second source of evidence comes from the randomized portion of the trial (i.e., third group above). If patients in the experimental arm of have superior outcomes to those in the control arm, there is also evidence of activity.

Assuming activity is identified, the challenge is to predict which patients are most likely to benefit. This is an important goal of the correlative science associated with many clinical trials.
Summary: Adaptive Design
There are numerous ways to modify the traditional randomized controlled trial design as it accrues. The examples presented here were used to introduce some key design concepts. They also provide a sense of the types of innovations that are possible, and how they might influence patients and research progress. The questions presented in the “Questions to Ask about Clinical Trials Section” are as applicable to adaptive designs as to traditional designs. Advantages of adaptive designs often come at the cost of increased complexity and opportunity for abuse. Thus, in evaluating adaptive designs it is useful to keep in mind the following points, and raise questions about them:
- All aspects of adaptive designs, including all adaptation rules, must be fully specified prior to starting the trial.
- While some aspects of the design may not entail randomization (e.g., the first stages of the patient preference and random discontinuation designs), virtually all sound designs include some randomization.
- Randomization does not require equal numbers of patients in each intervention arm; however, a random process must assign individual patients to intervention arms in the desired proportion.
- Although frequentists may use multi-stage designs, they exact a severe penalty on power that diminish the efficiency gained by adaptive designs.
- Continuous data monitoring and adaptation is natural to Bayesians since they are interested in continuously updating probabilities, rather than determining “absolute truth.”
V. Conclusions

This tutorial was developed to provide research advocates with a basic understanding of the scientific method and two alternative inferential processes used to establish evidence-based clinical practice from clinical trials. Given this understanding, research advocates should be able to contribute to this process. While not experts in science, statistics nor trial design, research advocates have a unique contribution to make because they focus on the whole patient experience, have a sense of urgency about making progress, and are not afraid to ask naïve questions. Finally, the tutorial described a paradigm shift (Figure 11) that is underway in clinical trials. Research advocates are encouraged to become more knowledgeable by asking questions and continuing to read about these topics. The recommended reading list that follows should assist in this endeavor.
VI. Acknowledgements

Funding for this tutorial was provided by the Department of Defense Center of Excellence Grant in Breast Cancer Research at Indiana University to the Research Advocacy Network Advocate Institute. Copies can be downloaded from the Research Advocacy Network website at www.researchadvocacy.com.

Reviewers:
We gratefully acknowledge the important role and dedication of significant time to review the tutorial:

• Connie Cirrincione, MS, Cancer and Leukemia Group B Staff Statistician
• Mary Elliott, Breast Cancer Advocate
• Susan Nathanson, PhD, Breast Cancer Advocate and former Executive Director, Y-ME National Breast Cancer Organization
• Anne O’Neill, MS, Eastern Cooperative Oncology Group
• Elda Railey, Co-Founder Research Advocacy Network
• Mary Lou Smith, JD, MBA, Co-Founder Research Advocacy Network

Appreciation is also expressed to the following for additional feedback and advice to the author:

• Dan Hayes, MD, University of Michigan
• Marion Perlmutter, PhD, University of Michigan
• Don Berry, PhD, MD Anderson Cancer Center

Research Advocacy Network
The Research Advocacy Network is a non-profit organization working to bring together all participants in the medical research process with the focus on education, support and connecting patient advocates with the research community to improve patient care. The Research Advocacy Network has started the Advocate Institute to educate and equip patient advocates with basic scientific knowledge and a better understanding of the research system. The Institute employs the newest technologies to facilitate knowledge transfer.

Jane Perlmutter, PhD
Jane Perlmutter is a consultant-evaluator for the Higher Learning Commission and a frequent facilitator for its Academic Quality Improvement Program Strategy Forums and other programs. She has numerous publications and a Distinguished Teaching Award. Her formal education includes Masters degrees in Educational Psychology and Computer Science, and a Ph.D. in Cognitive Psychology from the University of Massachusetts in Amherst. She also received an MBA from New York University’s Executive Program. Jane is a long-term breast cancer survivor and research advocate. She can be contacted at janep@gemini-grp.com.
VII. Recommended Readings

**Introductory Statistics**

**History of Statistics**

**Drug Development**


**Bayesian Statistics**


**Adaptive Designs**
### VIII. Glossary

<table>
<thead>
<tr>
<th>Concept</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha ((\alpha)) Error</td>
<td>In a test of a statistical hypothesis, the probability of rejecting the null hypothesis when it is true. Also called a type I error or false positive.</td>
</tr>
<tr>
<td>Adaptation Rule</td>
<td>Pre-specified rule that defines how an adaptive trial may be changed. Examples include changing the allocation of patients to treatment arms, changing the total number of patients that will be recruited, adding or deleting treatment arms, or stopping the trial early.</td>
</tr>
<tr>
<td>Adaptive Design</td>
<td>Multi-stage design in which some aspect of later stages of a trial depend, in a pre-defined way, upon what happens during earlier stages of the trial.</td>
</tr>
<tr>
<td>Adaptive Trial</td>
<td>Clinical trial that employs an adaptive design.</td>
</tr>
<tr>
<td>Analysis of Variance (ANOVA)</td>
<td>A statistical method to assess whether the amount of variation in a process is significant or caused by chance.</td>
</tr>
<tr>
<td>Arm</td>
<td>Any of the treatment groups in a clinical trial. Many randomized trials have two arms—one experimental and one control—but some have three or more “arms.” Some phase II trials have only one arm.</td>
</tr>
<tr>
<td>Balanced</td>
<td>In trials with two or more treatment arms, ensuring that all treatment arms have approximately the same proportion of patients with a given characteristic, for example, gender or race.</td>
</tr>
<tr>
<td>Bayesian Approach</td>
<td>A form of statistical reasoning that is based on continuously learning or updating the probabilities associated with an event. In particular, prior probabilities are modified in the light of data or empirical evidence in accordance with Bayes’ theorem to yield posterior probabilities, which may then be used as prior probabilities for further updating in the light of subsequent data. This increasingly popular method represents an alternative to the traditional (or frequentist probability) approach: whereas the latter attempts to establish confidence intervals around parameters, and/or falsify a-prior null-hypotheses, the Bayesian approach attempts to keep track of how a-prior expectations about some phenomenon of interest can be refined, and how observed data can be integrated with such a-prior beliefs, to arrive at updated posterior expectations about the phenomenon.</td>
</tr>
<tr>
<td>Beta ((\beta)) Error</td>
<td>In a test of a statistical hypothesis, the probability of failing to reject the null hypothesis when it is in fact false. Beta errors are also called type II errors, misses or false negatives.</td>
</tr>
</tbody>
</table>
### Bias
Bias in a sample is the presence or influence of any factor that causes the population or process being sampled to appear different from what it actually is. Bias is introduced into a sample when data are collected without regard to key factors.

### Biomarker
A characteristic (e.g., protein, gene) that is objectively measured and evaluated from a biological sample (e.g., tissue, blood) as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.

### Blinding
A procedure in which one or more parties to the trial are kept unaware of the treatment assignment. Double (or triple) blinding means that investigators and/or health care providers, in addition to patients, are unaware of the treatment assignments.

### Clinical Endpoint
An occurrence that measures the study hypothesis. It is often a characteristic or variable that reflects how a patient feels, functions, or survives, and used to measure whether a treatment is effective.
- **Primary**—what a trial is designed to assess
- **Secondary**—other endpoints of interest

### Clinical Trial
A type of research study that assesses medical questions in people. These studies often test new methods of screening, prevention, diagnosis, or treatment of a disease.

### Comparison Group; Control Intervention or Control Treatment
The treatment received by the comparison or control group, often a placebo. In cancer trials, the control intervention is usually the current standard of care.

### Confounding
In research design, the problem that arises when two or more causal variables, often an independent variable and an extraneous variable, are not properly controlled, so that their separate effects on the outcome measure cannot be disentangled.

### Comparison or Control Group or Arm
A group of patients who are not treated with the experimental intervention (e.g., no therapy, a different therapy, or a placebo) This group is compared to the group that receives the experimental intervention, to see if the experimental intervention is effective.

### Correlation
A statistical technique for determining the extent to which variations in the values of one variable are associated with variations in the value of another.
<table>
<thead>
<tr>
<th><strong>Correlative Science</strong></th>
<th>A type of study that uses as its primary explanatory variables obtained from the laboratory, such as various genetic, proteomic, biomarker data extracted from tumors.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crossover</strong></td>
<td>Allowing patients who do not respond to the treatment to which they were randomly assigned, to switch to the alternative treatment after some pre-specified amount of time.</td>
</tr>
<tr>
<td><strong>Demographics</strong></td>
<td>Having to do with the structure of populations or population statistics (e.g., age, income, marital status).</td>
</tr>
<tr>
<td><strong>Dependent Variable</strong></td>
<td>A variable that is acted on or influenced by another variable. For example, in an investigation of the affect of drugs on cancer, the independent variable (e.g., drug type or dose) is manipulated and the affect of this manipulation can be seen in the change in the dependent variable or outcome (e.g., survival).</td>
</tr>
<tr>
<td><strong>Efficacy</strong></td>
<td>A drug’s ability to produce beneficial effects on the course or duration of a disease. Efficacy can be measured in several ways, depending on the type of clinical trial.</td>
</tr>
<tr>
<td><strong>Eligibility Requirements</strong></td>
<td>Specifications of who may participate in a clinical trial. Eligibility requirements generally include details about the disease (e.g., organ site, stage), prior treatment, and co-morbidities.</td>
</tr>
<tr>
<td><strong>Endpoint</strong></td>
<td>In clinical trials, an event or outcome that can be measured objectively to determine whether the intervention being studied is beneficial. For example, a clinical trial studying a new cancer drug might use death as an endpoint to determine if people getting the drug lived longer than those who did not get the drug.</td>
</tr>
<tr>
<td><strong>Equivalence Trial</strong></td>
<td>A clinical trial designed to evaluate whether an experimental treatment is similar to a control treatment, by an appropriate definition of similarity. A two-sided (two-tailed) test of similarity is used.</td>
</tr>
<tr>
<td><strong>Evidence-Based Medicine</strong></td>
<td>Evidence-based medicine, defined by David Sackett, is 'the conscientious and judicious use of current best evidence from clinical care research, in the management of individual patients.”</td>
</tr>
<tr>
<td><strong>Experiment</strong></td>
<td>A scientific procedure undertaken to make a discovery, test a hypothesis, or demonstrate a known fact.</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Experimental Design</strong></td>
<td>The general plan of an experiment, including the method of assigning research participants or patients to treatment conditions, controlling extraneous variables, manipulating the independent variable, and measuring the dependent variable or outcome.</td>
</tr>
<tr>
<td><strong>Experimental or Investigational Group or Arm</strong></td>
<td>In clinical trials, a group of patients who receive the experimental or investigational intervention.</td>
</tr>
<tr>
<td><strong>Experimental or Investigational Intervention</strong></td>
<td>Term often used to denote a therapy (drug, drug dose or combination, device, or procedure) that is unproven or not yet scientifically validated with respect to safety and efficacy in humans.</td>
</tr>
<tr>
<td><strong>Frequentist or Traditional Statistics</strong></td>
<td>Traditional approach to statistical inference. Using hypothesis testing, it allows researchers to draw inferences about how likely they are to observe their data (or more extreme data) if the null hypothesis were true. Limits the relative frequency of drawing erroneous conclusions. It does not, however, allow researchers to assess the relative likelihood of competing hypotheses (other than the null hypothesis) in light of their data.</td>
</tr>
<tr>
<td><strong>Historical Controls</strong></td>
<td>Control group based on previous trials. Using historical controls is less costly in terms of time and money than including a randomized control group in a new trial. However, it typically introduces many unknown biases and is not generally acceptable for phase III trials.</td>
</tr>
<tr>
<td><strong>Hypothesis</strong></td>
<td>A tentative proposal made to explain certain observations or facts that require further investigation to be verified.</td>
</tr>
<tr>
<td><strong>Hypothesis Testing</strong></td>
<td>Hypothesis testing refers to the process of using statistical analysis to determine if the observed differences between two or more samples are due to random chance (as stated in the null hypothesis) or to true differences. A null hypothesis (H0) is a stated assumption that there is no difference in outcomes for two or more populations. The alternate hypothesis is a statement that the observed difference or relationship between two populations is real and not the result of chance or an error in sampling. Hypothesis testing is the process of using a variety of statistical tools to analyze data and, ultimately, to reject or not reject the null hypothesis.</td>
</tr>
<tr>
<td><strong>Incidence</strong></td>
<td>The frequency of occurrence or onset of new cases of a disorder as a proportion of a population in a specific time period, usually expressed as the number of new cases per 100,000 per annum.</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Independent Variable</td>
<td>Variable that is manipulated, sometimes experimentally, in order to observe its effects on a dependent variable. For example, in an investigation of the affect of drugs on cancer, the independent variable (e.g., drug type or dose) is manipulated and the affect of this manipulation can be seen in the change in the dependent variable or outcome (e.g., survival). A variable that is acted on or influenced by another variable.</td>
</tr>
<tr>
<td>Inductive Methods</td>
<td>A form of reasoning, also called empirical induction, in which a general law or principle is inferred from particular instances that have been observed.</td>
</tr>
<tr>
<td>Inference</td>
<td>A conclusion reached on the basis of evidence and reasoning.</td>
</tr>
<tr>
<td>Informed Consent</td>
<td>A process in which a person is given important facts about a medical procedure or treatment, a clinical trial, or genetic testing before deciding whether or not to participate. It also includes informing the patient when there is new information that may affect his or her decision to continue. Informed consent includes information about the possible risks, benefits, and limits of the procedure, treatment, trial, or genetic testing.</td>
</tr>
<tr>
<td>Institutional Review Board (IRB)</td>
<td>An IRB is a committee with federal regulatory authority to review and approve research involving human subjects. An IRB is composed of a diverse group of men and women with expertise in science, ethics, and other non-scientific areas.</td>
</tr>
<tr>
<td>Intent-to-Treat</td>
<td>In reality, not all patients who enroll in a clinical trial complete the trial as planned. They may drop out, die, switch treatments, etc. How to deal with the data of such patients is problematic because the patients who drop out are often different from those who complete the trial. Thus, biostatisticians often conduct two analyses—one including all patients assigned to treatment arms (i.e., those intended-to-be-treated) and a second including only patients who actually were treated. When the results of these two analyses differ, it is likely that the intervention influenced the propensity of patients to drop out of the trial.</td>
</tr>
<tr>
<td>Interaction</td>
<td>The situation in which a treatment difference (e.g., difference between experimental and control) is dependent on another factor (e.g., study site, organ site, gender).</td>
</tr>
<tr>
<td><strong>Term</strong></td>
<td><strong>Definition</strong></td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Intervention</td>
<td>The act or instance of intervening. In a clinical trial, an experimental or investigational intervention is compared to a comparison or control intervention. The interventions are often different treatment drugs, but may simple entail different schedules of drug administration, supportive therapies, etc.</td>
</tr>
<tr>
<td>Investigational Treatment</td>
<td>A new drug, biological drug, or combination that is used in a clinical investigation.</td>
</tr>
<tr>
<td>In Vitro</td>
<td>In the laboratory (outside the body). The opposite of in vivo (in the body).</td>
</tr>
<tr>
<td>In Vivo</td>
<td>In the body. The opposite of in vitro (outside the body or in the laboratory). Applies to both human and other animal bodies.</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>In Bayesian inference, the ratio between the probability of an observation or datum conditional on a hypothesis (numerator) and the probability of the same observation or datum conditional on an alternative hypothesis (denominator).</td>
</tr>
<tr>
<td>Mean</td>
<td>Measure of central tendency. Equally weighs all outcomes, and can be heavily influenced by outliers. Computationally, the mean is equal to the sum of outcomes, divided by the sample size.</td>
</tr>
<tr>
<td>Median</td>
<td>Measure of central tendency in which half of the outcomes are above and half below. The median is not affected by outliers.</td>
</tr>
<tr>
<td>Meta-analysis</td>
<td>The process of combining the data from a number of independent studies (usually drawn from published literature) and synthesizing summaries and conclusions addressing a particular issue. It aims to utilize the increased power of pooled data to clarify the state of knowledge on that issue. Meta analysis is often used in systematic reviews of studies of medical therapies to evaluate therapeutic effectiveness.</td>
</tr>
<tr>
<td>Multi-Stage Trial Design</td>
<td>Trials in which later stages of the trial are dependent upon what happens in earlier stages. An example of a two-stage trial design is the randomized discontinuation design where a patient’s treatment may be changed in the second stage of the trial, depending on the progress of the disease during the first stage.</td>
</tr>
<tr>
<td>Non-inferiority Trial</td>
<td>A clinical trial designed to evaluate whether an experimental treatment is similar to a control treatment, by an appropriate definition of similarity. A one-sided (one-tailed) test of similarity is used.</td>
</tr>
<tr>
<td>Null Hypothesis (H₀)</td>
<td>A null hypothesis (H₀) is a stated assumption that there is no difference in outcomes for two or more populations. According to the null hypothesis, any observed difference in samples is due to chance or sampling error. The term that statisticians often use to indicate the statistical hypothesis being tested.</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>One-Tail Test</td>
<td>Test for deviation from the null hypothesis in only one direction. That is, the null hypothesis states that a specific group is superior to the other.</td>
</tr>
<tr>
<td>Opportunity Costs</td>
<td>The economic cost of an action measured in terms of the benefit foregone by not pursuing the best alternative course of action. In clinical research, opportunity costs are measured relative to alternative trials that could have been conducted on the same patient population, by the same researchers, and/or with the same funds.</td>
</tr>
<tr>
<td>Outcome</td>
<td>In clinical trials, an event or outcome that can be measured objectively to determine whether the intervention being studied is beneficial. In cancer clinical trials outcomes that are of common interest include overall survival time, disease-free survival time, five year survival rate, and proportion of patients’ disease who respond to treatment.</td>
</tr>
<tr>
<td>p Value</td>
<td>The lowest level of significance at which a given null hypothesis can be rejected; that is, the probability of observing a result as extreme as or more extreme than that observed if the null hypothesis were true.</td>
</tr>
<tr>
<td>Paradigm Shift</td>
<td>A fundamental change in approach or underlying assumptions.</td>
</tr>
<tr>
<td>Parameter</td>
<td>A population parameter is the value of some quantitative characteristic in an entire population. It is estimated by a sample statistic.</td>
</tr>
<tr>
<td>Patient Allocation</td>
<td>Adaptive design in which the proportion of patients assigned to each treatment arm is modified as data are accrued. Compared to traditional designs, these design generally come to conclusions faster and require fewer patients, and with a larger proportion of the patients receiving the superior treatment.</td>
</tr>
<tr>
<td>Patient Enrichment</td>
<td>In the context of clinical trials, patient enrichment entails restricting patient eligibility to those most likely to benefit from the experimental treatment. As cancer treatments become more targeted, patients who are most likely to benefit can be selected based on biomarkers in the tumor or circulating cells, proteins in the blood, or genetic factors that influence the way drugs are metabolized.</td>
</tr>
<tr>
<td>Category</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Patient Preference Design</td>
<td>Design in which patients decide whether or not to be randomized, or to select their own treatment. Compared to traditional designs, these designs generally accrue more rapidly and are more satisfactory to patients.</td>
</tr>
<tr>
<td>Phase 0 Trial</td>
<td>Phase 0 trials are a novel concept in clinical trials. They involve testing small, non-therapeutic amounts of drug to obtain preliminary pharmacokinetic information, and assist pharmaceutical companies in decisions on pursuing further development of the agent. Pharmacokinetics is the study of the metabolism and action of drugs with particular emphasis on the time required for absorption, duration of action, distribution in the body, and method of excretion.</td>
</tr>
<tr>
<td>Phase I Trial</td>
<td>The first step in testing a new treatment in humans. These studies test the best way to give a new treatment (for example, by mouth, intravenous infusion, or injection) and the highest tolerable dose. The dose is usually increased a little at a time in order to find the highest dose that does not cause harmful side effects. Because little is known about the possible risks and benefits of the treatments being tested, Phase I trials usually include only a small number of patients who have not been helped by other treatments without a comparison group.</td>
</tr>
<tr>
<td>Phase II Trial</td>
<td>A study to test whether an experimental intervention has an anticancer effect (for example, whether it shrinks a tumor or improves blood test results) and whether it works against a certain type of cancer.</td>
</tr>
<tr>
<td>Phase III Trial</td>
<td>A study to compare the results of people taking an experimental intervention with the results of people taking the standard of care (for example, which group has better survival rates or fewer side effects). In most cases, studies move into phase III only after an intervention seems to work in phases I and II. Phase III trials may include hundreds of people and always includes a control group.</td>
</tr>
<tr>
<td>Phase IV Trial</td>
<td>A study conducted after a treatment has been approved and is being marketed to evaluate side effects that were not apparent in the phase III trial. Thousands of people are involved in a phase IV trial.</td>
</tr>
</tbody>
</table>
| PICO              | The acronym PICO is used by health professionals to convey all elements of the clinical scenario in an orderly fashion:  

  - **P** - patient, population of patients, problem  
  - **I** - intervention (a therapy, test)  
  - **C** - comparison (another therapy, placebo)  
  - **O** - outcome (survival, response)
<table>
<thead>
<tr>
<th><strong>Pivotal Trial</strong></th>
<th>A controlled trial to evaluate the safety and efficacy of a drug in patients who have the disease or condition to be treated.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Placebo</strong></td>
<td>An inactive substance or treatment that looks the same as, and is given the same way as, an active drug or treatment being tested. The effects of the active drug or treatment are compared to the effects of the placebo.</td>
</tr>
<tr>
<td><strong>Population</strong></td>
<td>The entire collection of people (current and future) who are the focus of interest (e.g., all people with a specific type of cancer).</td>
</tr>
<tr>
<td><strong>Population Parameter</strong></td>
<td>A value of some quantitative characteristic in a population. Population parameters are estimated by sample statistics calculated from sample data (e.g., sample mean).</td>
</tr>
<tr>
<td><strong>Posterior Probability</strong></td>
<td>The posterior probability is the conditional probability of a variable, taking the evidence into account. The posterior probability is computed from the prior probability and the likelihood function via Bayes’ theorem.</td>
</tr>
<tr>
<td><strong>Power</strong></td>
<td>Power is the probability of rejecting the null hypothesis if it is really false. It is mathematically equal to $1 - b$ and is dependent upon the sample size, sample variance, the effect size and type II error rate.</td>
</tr>
<tr>
<td><strong>Prevalence</strong></td>
<td>The total number of existing cases of a disorder as a proportion of a population (usually per 100,000 people) at a specific time.</td>
</tr>
<tr>
<td><strong>Primary Endpoint</strong></td>
<td>The main result that is measured to see if a given treatment worked (e.g., the number of deaths or the difference in survival between the treatment group and the control group). The primary endpoint is determined by the primary study objective, and is defined prior to the start of the trial.</td>
</tr>
<tr>
<td><strong>Prior Probability</strong></td>
<td>A prior probability is a base rate, interpreted as a description of what is known about a variable in the absence of some evidence.</td>
</tr>
<tr>
<td><strong>Probability</strong></td>
<td>The likelihood that a given event will occur. Probability is expressed as values between 0 (complete certainty that an event will not occur) to 1 (complete certainty that an event will occur), or percentage values between 0 and 100%.</td>
</tr>
<tr>
<td><strong>Protocol</strong></td>
<td>An action plan for a clinical trial. The plan states what the study will do, how, and why. It explains how many people will be in it, who is eligible to participate, what study agents or other interventions they will be given, what tests they will receive and how often, and what information will be gathered.</td>
</tr>
<tr>
<td><strong>Quality of Life (QOL)</strong></td>
<td>Measurement of aspects of an individual’s sense of well-being and ability to perform various tasks.</td>
</tr>
<tr>
<td><strong>Randomization</strong></td>
<td>The process by which patients are assigned by chance to separate groups that compare different treatments or other interventions. Randomization can use equal weighting (i.e., 50:50) or not (e.g., 75:25)</td>
</tr>
<tr>
<td><strong>Randomized Controlled Trial</strong></td>
<td>A research design used for testing the effectiveness of a drug, or any other type of experimental intervention, in which research participants are assigned randomly to experimental and control or groups and the differences in outcomes are compared.</td>
</tr>
<tr>
<td><strong>Randomized Discontinuation Design</strong></td>
<td>Design in which all patients initially receive the experimental treatment. In a second stage of the trial a subgroup of patients is randomized. Compared to traditional designs, these designs generally provide better information about the sub-set of patients most likely to benefit from an experimental treatment and are often preferred by patients.</td>
</tr>
<tr>
<td><strong>Reliability</strong></td>
<td>Reliability is the extent to which an experiment, test, or any measuring procedure yields the same result on repeated trials. Without the agreement of independent observers able to replicate research procedures, or the ability to use research tools and procedures that yield consistent measurements, researchers would be unable to satisfactorily draw conclusions, formulate theories, or make claims about the generalizability of their research.</td>
</tr>
<tr>
<td><strong>Research Advocate</strong></td>
<td>Individuals or organizations who try to raise public awareness about important cancer issues, particularly those related to research. They work with researchers to ensure that research is patient focused and likely to result in changes in clinical practice.</td>
</tr>
<tr>
<td><strong>Sample</strong></td>
<td>A subset of a population selected to draw inferences about the population. It is a random sample if it is chosen in such a way that every sample of the same size has an equal chance of being selected.</td>
</tr>
<tr>
<td><strong>Sampling Distribution</strong></td>
<td>Every statistic is a random variable because its value varies from one sample to another. The distribution of this random variable is a sampling distribution.</td>
</tr>
<tr>
<td><strong>Scientific Method</strong></td>
<td>A method of procedure that has characterized natural science since the 17th century, consisting of systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses.</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Secondary Endpoints</strong></td>
<td>These are outcomes that are of interest in addition to the primary endpoints that a clinical trial is designed to assess. Examples include quality of life (QOL) measures and treatment side effects.</td>
</tr>
<tr>
<td><strong>Sensitivity</strong></td>
<td>The conditional probability of a test correctly giving a positive result, given that the patient does have the disease.</td>
</tr>
<tr>
<td><strong>Side Effect</strong></td>
<td>A problem that occurs when treatment affects healthy tissues or organs. Some common side effects of cancer treatment are fatigue, pain, nausea, vomiting, decreased blood cell counts, hair loss, and mouth sores. Serious side effects are often referred to as adverse events.</td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td>The conditional probability of a test correctly giving a negative result, given that the patient does not have the disease.</td>
</tr>
<tr>
<td><strong>Standard of Care or Standard Treatment</strong></td>
<td>In medicine, treatment that experts agree is appropriate, accepted, and widely used. Health care providers are obligated to provide patients with the standard of care. Also called standard therapy or best practice.</td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>Measure of dispersion calculated from samples and used to estimate population variances. Computationally, the standard deviation is equal to the square root of the variance.</td>
</tr>
<tr>
<td><strong>Statistic</strong></td>
<td>A statistic is the value of some quantitative characteristic in a sample taken to be an estimate of the equivalent population parameter.</td>
</tr>
<tr>
<td><strong>Statistics</strong></td>
<td>The scientific discipline concerned with the collection, analysis, interpretation, and presentation of data.</td>
</tr>
<tr>
<td><strong>Statistical Analysis</strong></td>
<td>Statistical analysis relates observed statistical data to theoretical models, such as probability distributions or models used in regression analysis. By estimating parameters in the proposed model and testing hypotheses about rival models, one can assess the value of the information collected and the extent to which the information can be applied to similar situations. Statistical prediction is the application of the model thought to be most appropriate, using the estimated values of the parameters.</td>
</tr>
<tr>
<td><strong>Statistical Inference</strong></td>
<td>Statistical inference involves the selection of one conclusion from a number of alternatives according to the result of a calculation based on observations.</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Statistical Significance</strong></td>
<td>A term indicating that the results of a study are stronger than would be expected by chance alone.</td>
</tr>
<tr>
<td><strong>Stratification</strong></td>
<td>The placing of the trial population into categories (i.e., strata) which are: 1) exhaustive (i.e., all strata together include the entire trial population); 2) mutually exclusive and; 3) related to the criteria being studied.</td>
</tr>
<tr>
<td><strong>Stratification Variable</strong></td>
<td>The variable which form the basis of stratification. Examples include gender or age, disease site or stage.</td>
</tr>
<tr>
<td><strong>Sub-Group Analysis</strong></td>
<td>Analyses that look for treatment difference in sub-groups of the experimental and control groups. For example, is there a treatment effect in males but not females or in one organ site but not another. If these analyses are planned prior to running the trial, statistical procedures can be used to limit a or type I errors, although at a cost to power. They can be especially problematic when they are not pre-planned.</td>
</tr>
<tr>
<td><strong>Superiority Trial</strong></td>
<td>A clinical trial designed to evaluate whether an experimental treatment is superior to a control treatment, by an appropriate definition of similarity. A one-sided test would be used.</td>
</tr>
<tr>
<td><strong>Supportive Therapy</strong></td>
<td>A treatment designed to improve, reinforce, or sustain a patient’s physiological well-being or psychological self-esteem and self-reliance. In cancer trials, supportive therapies are typically given to prevent or minimize toxic side-effects of therapies used to treat the cancer.</td>
</tr>
<tr>
<td><strong>Surrogate Endpoint</strong></td>
<td>A biomarker that is intended to substitute for a clinical endpoint (e.g., survival). A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence.</td>
</tr>
<tr>
<td><strong>Test Statistic</strong></td>
<td>A statistic used in hypothesis testing. It has a known distribution if the null hypothesis is true.</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td>Any intervention—drug, surgery, psychosocial intervention—being investigated in a clinical trial.</td>
</tr>
<tr>
<td><strong>Treatment Arm</strong></td>
<td>Any of the treatment groups in a randomized trial. Many randomized trials have two arms—one experimental and one control—but some have three or more “arms,” and some have only one.</td>
</tr>
<tr>
<td><strong>Treatment Effect</strong></td>
<td>The treatment effect is the difference in outcomes between the group of patients who received the experimental or investigational treatment and those who received the comparison or control treatment.</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Treatment Protocol</strong></td>
<td>The treatment protocol specifies all details about the interventions that are to be given to both the experimental or investigational group and the comparison or control group. This includes drug dose and schedule, as well as test procedures and any other interventions that are part of the trial.</td>
</tr>
<tr>
<td><strong>Two Tail Test</strong></td>
<td>Testing for deviation from the null hypothesis in either direction.</td>
</tr>
<tr>
<td><strong>Type I Error</strong></td>
<td>In a test of a statistical hypothesis, the probability of rejecting the null hypothesis when it is true. Also called an alpha error or false positive.</td>
</tr>
<tr>
<td><strong>Type II Error</strong></td>
<td>In a test of a statistical hypothesis, the probability of failing to reject the null hypothesis when it is in fact false. Type II errors are also called beta errors, misses or false negatives.</td>
</tr>
<tr>
<td><strong>Underserved Patient Populations</strong></td>
<td>Populations whose participation in clinical trials is less than their representation in the overall population of people affected by a disease. Underserved populations typically include minorities, people of lower socio-economic background, the elderly, and people who live in rural areas.</td>
</tr>
<tr>
<td><strong>Validity</strong></td>
<td>Validity refers to the degree to which a study accurately reflects or assesses the specific concept that the researcher is attempting to measure. Researchers should be concerned with both external and internal validity. External validity refers to the extent to which the results of a study are generalizable or transferable. Internal validity refers to: 1) the rigor with which the study was conducted (e.g., the study’s design, the care taken to conduct measurements, and decisions concerning what was and wasn’t measured); and 2) the extent to which the designers of a study have taken into account alternative explanations for any causal relationships they explore.</td>
</tr>
<tr>
<td><strong>Variable</strong></td>
<td>The characteristic measured or observed when an experiment is carried out or an observation is made. Variables may be non-numerical or numerical. Since a non-numerical observation can always be coded numerically, a variable is usually taken to be numerical. Statistics is concerned with random variables and with variables whose measurement may involve random errors.</td>
</tr>
<tr>
<td>Variability</td>
<td>The degree to which a set of outcomes is dispersed or scattered. Two sets of outcomes with identical means (averages) may have widely different variances. The usual measures of variability are the variance and standard deviation.</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Variance</td>
<td>Measure of dispersion calculated in samples and used to estimate population variances. Computationally, the variance is equal to the average squared deviation from the mean.</td>
</tr>
</tbody>
</table>
Glossary Sources

http://www.cancer.gov/dictionary/
http://www.cancer.gov/statistics/glossary
http://www.fda.gov/cder/drugsatfda/glossary.htm
http://www.med.umich.edu/cacr/dictionary/A-B.htm
http://www.isixsigma.com/dictionary/
http://writing.colostate.edu/guides/research/relval/