OTC FOUNDATION
CLINICAL RESEARCH COURSE

Practical Tips in the Design and Conduct of Research

Generic One-Day Course Outline

DAY 1 -MORNING SESSION		
Time	**Lecture**	**Speaker**
Session I	**Principles of Evidence-Based Orthopaedics**	
9:00am	Why is Clinical Research Necessary for Surgeons?	
9:15am	What is Evidence-Based Orthopedics?	
9:30am	What is the Language of EBM?	
9:45am	Discussion	
Session II	**Study Designs**	
10:00 am	Hierarchy of Research Design?	
10:15am	What is a Clinical Case Series?	
10:30am	What is a Case-Control Study?	
10:45am	What is a Prospective Cohort Study?	
11:00 am	Discussion	
11:15 am	BREAK/REFRESHMENTS/SNACK	
11:30 am	What is a Randomized Trial?	
11:45 am	What is a Systematic Review (Meta-analysis)?	
12:00 pm	How to Limit Bias in Clinical Research?	
12:15 am	Panel Discussion	
12:30am-	LUNCH	

DAY 1 -AFTERNOON SESSION		
Time	**Lecture**	**Speaker**
Session III	**Doing Research**	
1:30 pm	How to prepare a research proposal?	
1:45 pm	What infrastructure is needed to conduct a clinical study?	
2:00 pm	How many patients will I need?	
2:15 pm	Multicenter research: How can I participate?	
2:30 pm	How to set up an ethics approval for your study	
2:45 pm	Discussion	
3:00 pm	BREAK/REFRESHMENTS	
3:15pm	What outcome measure should I use?	
3:30 pm	Basic Statistics: P value and 95% Confidence interval	
3:45 pm	Writing your Paper for Publication-Do's and Don'ts	
4:00 pm	Getting Involved: The potential for Cuba to lead clinical research	
4:15 pm	Discussion	
4:30 pm	Adjourn	

Evidence-Based Orthopedics

History of Evidence-Based Medicine

When Gordon Guyatt first introduced the term evidence-based medicine (EBM) in an informal residency training program document, he illustrated it as "an attitude of enlightened skepticism toward the application of diagnostic, therapeutic, and prognostic technologies in their day-to-day management of patients". He later described EBM in an editorial and the term as well as the philosophy – fundamentally influenced by the early work of clinical epidemiologists at McMaster University headed by David Sackett - became widespread through the series of articles published by the evidence-based medicine working group.

Evidence-based medicine's evolution has expanded to include evidence-based nursing, physiotherapy, occupational therapy, podiatry, and specialization. We need evidence-based obstetrics, gynaecology, internal medicine and surgery -- and of course, we need evidence-based orthopaedic surgery (EBO).

How Evidence-Based Orthopaedics Differs from the Traditional Approaches to Health Care Provision

According to the traditional paradigm, clinicians evaluate and solve clinical problems by reflecting on their own clinical experience or the underlying biology and pathophysiology or by consulting a textbook or local expert. For many traditional practitioners, reading the Introduction and Discussion sections of a research article is adequate for gaining relevant information, and observations from day-to-day clinical experience are a valid means of building and maintaining knowledge about patient prognosis, the value of diagnostic tests, and the efficacy of treatment. The study and understanding of basic mechanisms of disease and pathophysiologic principles are thus adequate guides for clinical practice. Because this paradigm places high value on traditional scientific authority and adherence to standard approaches, traditional medical training and common sense are sufficient bases for evaluating new tests and treatments, and content expertise and clinical experience are sufficient bases from which to generate guidelines for clinical practice.

While pathophysiology, clinical experience, and intuition are necessary, they, alone, are insufficient guides for practice because they may be incorrect and lead to inaccurate predictions about the performance of diagnostic tests and the efficacy of treatments. Like the traditional approach to health care, the evidence-based health care paradigm also assumes that clinical experience and the development of clinical instincts (particularly with respect to diagnosis) are crucial elements of physician competence. However, the EBO approach includes several additional steps. These steps include using your experience to identify your important knowledge gaps and information needs, phrasing those needs in the form of answerable questions, identifying potentially relevant research, accurately assessing the validity of evidence and results, developing clinical policies that align research evidence and clinical circumstances, and appropriately applying research evidence to individual patients with their particular experiences, expectations, and values.

Evidence-based medicine practitioners (i.e., clinicians who work under the EBO paradigm) regularly consult original literature, including the methods and results sections of research articles, and other reliable sources of evidence to help them solve clinical problems. Effectively using the literature requires an ability to independently assess the evidence and the credibility of the expert opinion offered. In turn, correctly interpreting literature on prognosis, diagnostic tests, and treatment and potentially harmful exposures (side effects, environmental exposures) requires an understanding hierarchy of evidence. For example, in making treatment decisions, EBO practitioners may conduct an N of 1 randomized trial (randomized trial in an individual patients, with patient repeatedly treated with active intervention or placebo) to determine the optimal treatment for an individual patient, or they may seek a systematic review of randomized trials of treatment alternatives. If a systematic review is not available, they will look for individual randomized trials and high quality observational studies of relevant management strategies. If the literature is altogether barren, EBO practitioners will fall back on the underlying biology and pathophysiology and clinical experience.

Summary

While EBO is sometimes perceived as a blinkered adherence to randomized trials it more accurately involves informed and effective use of all types of evidence, but particularly evidence from the medical literature, in patient care. With the ever-increasing amount of available information, surgeons must consider a shift in paradigm from traditional practice to one which involves question formulation, validity assessment of available studies and appropriate application of research evidence to individual patients.

Hierarchy of Research Design

Introduction

This section provides an approach to organizing published research on the basis of study design, a hierarchy of evidence. The key features and the advantages and disadvantages of specific study designs will be addressed. The concepts presented hopefully will enable clinicians and healthcare personnel to practice in the context of evidence-based orthopaedics.

Types of Study Design

The types of study designs used in clinical research can be classified broadly according to whether the study focuses on describing the distributions or characteristics of a disease or elucidating its determinants.

Descriptive studies describe the distribution of a disease, particularly what type of people have disease, in what locations and when. Cross sectional studies, case reports, and case series represent types of descriptive studies.

Case reports are an uncontrolled, descriptive study design involving an intervention and outcome with a detailed profile of one patient. Expansion of the individual case report to include multiple patients is a case series. Although descriptive studies are limited in their ability to make causal inferences about the relationship between risk factors and an outcome of interest, they are helpful in developing a hypothesis that can be tested using an analytic study design.

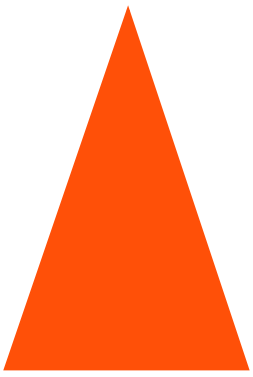
Analytic studies focus on determinants of a disease by testing a hypothesis with the ultimate goal of judging whether a particular exposure causes or prevents disease. Analytic design strategies are broken into two types**: observational studies**, such as case-control and cohort studies, and **experimental studies**, also called clinical trials. The difference between the two types of analytic studies is the role that the investigator plays in each of the studies. In the observational study, the investigator simply observes the natural course of events. In the trial, the investigator assigns the intervention or treatment.

One type of observational study is the case-control study that starts with the identification of individuals who already have the outcome of interest, cases, and are compared with a suitable control group without the outcome event. The relationship between a particular intervention or prognostic factor and the outcome of interest is examined by comparing the number of individuals with each intervention or prognostic factor in the cases and controls. Case-control studies are described in greater detail later.

In the **cohort study design**, the cohort represents a group of people followed over time to see whether an outcome of interest develops. Ideally this group meets a level of certain predetermined criteria representative of a population of interest and is followed with well-defined outcome variables. Usually this group is matched with a control population selected on the presence or absence of exposure to a factor of interest. The purpose of this type of study is to describe the occurrence of certain outcomes with time and to analyze associations between prognostic factors and those outcomes. The prospective cohort study is described in greater detail later.

Randomized controlled trials classically are held as the standard against which all other designs should be measured. In a randomized controlled trial, patients are assigned to a treatment group or a control group. The control group usually receives an accepted treatment or no treatment at all, whereas the treatment group is assigned the intervention of interest. Randomized controlled trials are thought to represent the highest quality of evidence based on their methodological strengths of randomization of patient assignment and blinding of intervention and outcome. The randomized controlled trial is described in greater detail later.

Table: Hierarchy of Evidence

STUDY DESIGN	LESS BIAS
Randomized Controlled Trials	
Controlled Trials (e.g., controlled before-after studies)	
Case Control Studies and Cohort Studies	
Cross-sectional Studies	
Expert Opinion, Case Reports and Case Series	

Levels of Evidence

Investigators have attempted to minimize potential harm to patients by basing clinical decisions of the sorts of evidence that are least likely to be wrong. Two studies defined what was thought to be the evidence providing the least biased estimate of the effect of an intervention: a systematic review (or meta-analysis) documenting homogeneity in the results of a large number of high-quality randomized controlled trials (randomized blinded, complete follow-up, and analysis of all of those randomized, so called intention-to-treat analysis). Such evidence was termed, Level 1 Evidence. Investigators additionally categorized studies of an intervention based on an increasing degree of potential bias: systematic reviews with randomized controlled trials that reveal differences in treatment effect (heterogeneity); individual high-quality randomized controlled trials (Level IB evidence); less rigorous randomized controlled trials; cohort or observational studies (Level 2 evidence); case-control studies (Level 3 evidence); case series (Level 4 evidence) and then expert opinion (Level 5 evidence). Based on the various levels of evidence on a particular treatment, grades of recommendation can be determined. For example, Sackett has proposed the following grades of recommendations: 1) grade A – consistent Level 1 studies, 2) grade B – consistent Level 2 or Level 3 studies, 3) grade C – Level 4 studies, and 4) grade D – Level 5 evidence.

Table: Hierarchy of Evidence

Level of Evidence		**Grade of Recommendations**
Level I:	Large randomized trials with clear-cut results, and low risk of error, or Meta-analyses of Randomized Trials with homogenous (similar) study results and narrow confidence intervals	A
Level II:	Randomized trials with uncertain results and/or moderate to high risk of error; prospective cohort studies of high quality	B
Level III:	Case-control studies or meta-analyses of case-control studies	B
Level IV:	Case series with no controls	C
Level V:	Expert opinion without explicit critical appraisal, or based upon physiology or bench research	D

Systematic Reviews and Meta Analysis

The Conduct of a Systematic Review

The *sine qua non* of systematic reviews is the systematic search and identification of studies. Bhandari *et al* found that 83% of orthopedic meta-analyses explicitly stated the methods used to search for evidence and that 73% of the meta-analyses had reasonably comprehensive strategies. By systematic, we mean the use of a search protocol that lists all potential data sources and multiple (and frequently overlapping) strategies to consult them. The main concern is that important studies could be missed if reviewers only consult a narrow choice of data sources. Electronic databases, such as MEDLINE (a database produced by the U.S. National Library of Medicine and freely accessible), provide the bulk of data for most systematic reviews. MEDLINE indexes a large number of publications, mostly from the United States (EMBASE,

another popular electronic database, exhibits some important overlap with MEDLINE but has greater European and multidisciplinary coverage). For instance, the choice of searching just in MEDLINE also entails searching mostly for studies published in English. Limited searches (such as searching only for studies published in peer-reviewed journals, in English, or very recently) may lead to a biased sample of the available evidence (rather than to all the available evidence) and bias the evidence summary. One form of such bias may result from *publication bias*.

The next step in the conduct of a systematic review is to select studies for inclusion in the review. The question usually defines the type of study to include and therefore the inclusion and exclusion criteria. In defining these, as in all previous steps, subject matter experts and research methodologists usually have important input. Inclusion criteria tend to define the study designs of interest as well as the type of patients, interventions, control interventions, and outcomes that matter to answer the review question. These criteria tend to be broad and inclusive so that the results will be generalizable while still reflecting the focused nature of the question. The exclusion criteria tend to be few and to limit the sample of studies to increase the homogeneity and maintain the focus of the review. Reviewers should decide and define the inclusion and exclusion criteria in the review protocol. Bhandari *et al* showed that 78% of orthopedic meta-analyses explicitly reported their criteria for inclusion of studies into the review, but only 43% of them were judged as able to avoid bias in the selection of studies for review.

It is important to limit investigator bias throughout all aspects of the review. To reiterate the threats to validity, even with a protocol outlining the review methods, reviewing research is a retrospective exercise, and is thus subject to random as well as systematic error. The Cochrane Collaboration routinely submits review protocols for peer review and publishes protocols in the Cochrane Library for open and widespread criticism and commentary as another step to avoid bias in the review process. The application of inclusion and exclusion criteria to abstracts and papers in full text should be piloted to ensure that the reviewers apply them uniformly. Usually, two or more reviewers working independently apply these criteria to later quantify their disagreements (sometimes using statistical methods that measure agreement beyond chance like the *kappa* statistic). Substantial agreement between reviewers ($\kappa > 0.7$) indicates that criteria were clear, objective, and consistently applied. Disagreements are frequently resolved by discussion or through arbitration by a third reviewer. Demonstrating reliability of article selection does not eliminate any kind of error, but does decrease the probability of serious error. Similar safeguards should be in place when abstracting data from the studies. The data to be abstracted should be pertinent to the review question, be specified in the review protocol, and be abstracted systematically. A duplicate independent process should be used, similar to that described for article selection. Data abstraction is best done after a structured form has been developed and piloted by reviewers, to enhance the efficient and accurate abstraction of key information. After data abstractors quantify and resolve disagreements, data are entered into a database (sometimes in duplicate or with multiple verification strategies depending on the data volume and capacity of the review team).

Perhaps the most important element to abstract is the methodological quality of the primary studies. Two of the methodological features of randomized trials for which there is empirical evidence of bias include allocation concealment, and blinding. Allocation concealment, such that all involved in the study cannot determine the arm to which the next

patient enrolled will be allocated to, is infrequently reported. Blinding and which parties in the study remained unaware of allocation after randomization and throughout the study also are poorly described. In a recent summary, poor methodological quality of the primary studies accounted for an overestimation of treatment effects by 15 due to inadequate blinding and by 30% due to inadequate concealment of the allocation sequence (in contrast with the impact of the reporting bias we discussed in the previous paragraphs which appears not to exceed 10%)[23]. In spite of its importance, reviewers may encounter challenges in abstracting the methodological quality elements from each study since many elements are poorly reported in the studies. Bhandari et al reported that 20% of trials provide enough information to ascertain the appropriateness of allocation concealment and that 11% reported on blinding of the outcome assessors (often the only practical blinding procedure in surgical trials). As in previous steps of the review requiring judgment, the provision of duplicate and independent assessments study methodology with quantification and resolution of disagreements by consensus or arbitration minimize the introduction of bias.

Once found, systematic reviews need to be critically appraised to determine the degree with which the methods used limit bias, the results and their precision, and the applicability of the results to the clinical situation that triggered the inquiry. This strategy of appraisal is included in the Users' Guides Series that have been recently summarized in a book by the members of the Evidence-based medicine working group and the applied appraisal guide to systematic reviews is summarized in the table below.

Table: Users' Guides for How to Use a Review Article

Are the results valid?
• Did the review explicitly address a sensible clinical question?
• Was the search for relevant studies detailed and exhaustive?
• Were the primary studies of high methodologic quality?
• Were assessments of studies reproducible?
What are the results?
• Were the results similar from study to study?
• What are the overall results of the review?
• How precise were the results?
How can I apply the results to patient care?
• How can I best interpret the results to apply them to the care of patients in my practice?
• Were all clinically important outcomes considered?
• Are the benefits worth the costs and potential risks?

Summary

Systematic reviews of the available evidence play a fundamental role in establishing what we know and do not know. They inform clinicians wanting to practice evidence-based medicine. Summaries of evidence also facilitate clinical practice by allowing it to be based on the available evidence and leaving the clinician and the patient with time to weight other elements into the decision such as the patient values and preferences for health care and the clinical circumstances. They are valuable teaching tools. Furthermore, most granting agencies demand a systematic

review of the literature to critically appraise what has been done before, to build on prior strengths, and create new and necessary knowledge to address residual uncertainty. Thus, systematic reviews and meta-analyses help to identify areas in need of further research. The systematic summary of the evidence is an expensive and labor-intensive research effort that cannot be realistically conducted by clinicians on the fly during clinical practice. Clinicians, teachers and researchers in the discipline of orthopedics should consider reading and conducting not only clinical trials of orthopedic interventions, but also systematic reviews of previously conducted research. We hope this discussion will further that enthusiasm.

Randomized Controlled Trials

In the hierarchy of research designs, the results of randomized controlled trials are considered the highest level of evidence. Randomization is the only method for controlling for known and unknown prognostic factors between two comparison groups. Lack of randomization predisposes a study to potentially important imbalances in baseline characteristics between two study groups. The role of nonrandomized (observational) studies in evaluating treatments is an area of continued debate: deliberate choice of the treatment for each patient implies that observed outcomes may be caused by differences among people being given the two treatments, rather than the treatments alone. Unrecognized confounding factors can interfere with attempts to correct for identified differences between groups. There has been considerable debate about whether the results of nonrandomized studies are consistent with the results of randomized controlled trials. Nonrandomized studies, or observational studies, have been reported to overestimate or underestimate treatment effects.

Investigators have attempted to minimize potentially harming patients by basing clinical decisions of the sorts of evidence that are least likely to be wrong. Two studies defined what was thought to be the evidence providing the least biased estimate of the effect of an intervention: a systematic review documenting homogeneity in results of a large number of high-quality randomized controlled trials (randomized with concealment, blinded, complete follow-up, and intention-to-treat analysis). This was termed, Level 1 Evidence, and the recommendations from this evidence was designated as Grade A. These considerations have supported a hierarchy of evidence, with randomized controlled trials at the top, controlled observational studies in the middle, and uncontrolled studies and opinion at the bottom. However, these findings have not been supported in two recent publications in the New England Journal of Medicine that identified non-significant differences in results between randomized, controlled trials and observational studies.

Randomized controlled trials classically are held as the standard to which all other designs should be measured. In a randomized controlled trial, patients are assigned to a treatment group or a control group. The control group usually receives an accepted treatment or no treatment at all, whereas the treatment group is assigned the intervention of interest. Randomized controlled trials are thought to represent the highest quality of evidence based on their methodological strengths of randomization of patient assignment and blinding of intervention and outcome. Studies are randomized to eliminate selection bias and to balance

confounding factors between both groups. Blinding of the subjects, investigators, or both (double-blinded) involves concealing patient assignment so as not to influence the outcome. Controlled trials without randomization occasionally are conducted but represent a class of evidence with less internal validity and subject to selection bias. One example of the importance of randomization comes from the neurosurgical literature. During the 1970s and early 1980s, surgeons frequently did extracranial to intracranial bypass (anastomosis of a branch of the external carotid artery, the superficial temporal, to a branch of the internal carotid artery, the middle cerebral). They thought it prevented strokes in patients whose symptomatic cerebrovascular disease otherwise was surgically inaccessible. Comparisons of outcomes among nonrandomized cohorts of patients, who, for various reasons, did or did not have this operation, fueled their conviction. These studies suggested that patients who had surgery seemed to fare much better. However, to the surgeons' surprise, a large multicenter trial in which patients were allocated to surgical or medical treatment using a process analogous to flipping a coin (a randomized control trial), showed that the only effect of surgery was to increase adverse outcomes in the immediate post-surgical period.

Another poignant example comes from the literature evaluating the role of ultrasound in fracture healing. A survey of physiotherapists identified a belief that the application of ultrasound was contraindicated in fractures. These perceptions were fueled by teaching in common textbooks. However, a systematic review of the literature to identify the current evidence contradicted the previous paradigm presented in most physiotherapy textbooks. When randomized trials were critically evaluated, and their results statistically pooled, low intensity ultrasound led to significantly shorter time to fracture healing, as measured by radiographs, when compared with placebo (mean effect size=64 days; 95% confidence interval = 10.1 to 118 days).

The advantages of a randomized controlled trial are the quality of the study associated with its inherent internal validity because potential confounding variables can be controlled for, thereby potentially providing strong evidence for cause and effect relationships. Randomized controlled trials may not always be suitable for answering some research questions for technical or ethical reasons. It is understandable that not all questions in surgery can be addressed by a randomized controlled trial; however, the potentially important information derived from such studies in the current climate of evidence-based orthopaedics is a compelling argument in their favor.

Randomized controlled trials are not appropriate or feasible for all surgical interventions, and it has been estimated that nearly 60% of surgical research questions could not be answered by an RCT, even in ideal clinical circumstances. Furthermore, in cases where an RCT is appropriate there still exists a choice of trial design, each of which has strengths and limitations. RCTs can be classified according to: [1] the aspect of the interventions that investigators want to explore, [2] the way in which participants are exposed to interventions, and [3] the number of participants under study. There are four basic RCT designs, which are described below.

Table: A Comparison of Randomized Controlled Trial Designs

RCT Design	Advantages	Disadvantages
Parallel	- simple design that can be applied to most interventions and illnesses	- cannot provide patient-specific information - each intervention under study requires a large increase in sample size
Cross-Over	- smaller sample required - within subject comparisons ensure that all baseline characteristics are equally distributed	- vulnerable to carry-over effects and period effects - only possible to test rapid acting interventions on chronic conditions
Factorial	- allows for the effect of combined therapies to be assessed	- vulnerable to interaction effects
n-of-1	- provides patient-specific information	- results are not generalizable

Paralle1 Trial Design

Most RCTs make use of a parallel design, in which participants are randomized to two or more groups of equal size and each group is exposed to a different intervention. This trial design produces between participant comparisons. Parallel trials with more than one treatment arm (matched to a control arm) provide an opportunity to study multiple interventions or different exposures to an intervention; however, this also demands larger sample sizes in order to ensure such trials are adequately powered to detect clinically significant differences between interventions.

Cross-Over Trial Design

A cross-over trial design assures that each study participant will receive all study interventions; however, the order in which they receive the interventions is random. As such each participant acts as their own control. Cross-over trials produce within participant comparisons, and thus they require less participants than parallel trials to produce statistically and clinically significant results.

There are a number of criteria that must be met in order for a cross-over trial design to be appropriate: [1] Study participants must be afflicted with chronic or incurable conditions. Conditions that may be resolved after a single intervention would not be suitable for cross-over trials; [2] Study interventions should have a rapid onset and a short duration of effect. Interventions with a long duration of effect risk a "carry-over effect" – when the effect of an intervention persists during the testing of another intervention. When the duration of an intervention is known, treatment periods can be separated by sufficient time to allow for the effect to run its' course. This period of time between treatments is known as the "washout

period"; [3] The condition under study must be stable over time in order to assure that any effect noted during the study can be attributed to the treatment provided and not simply a change in the condition that would have occurred with or without treatment. Differences between study periods that are the result of fluctuations in the condition being studies, and not the result of an intervention, are known as "period effects".

Factorial Design Trials

An RCT using a factorial design allows interventions to be evaluated both individually and in combination with one another. Therefore, two or more hypotheses can be explored in one experiment and the effect of combination therapy can be assessed. For example, in an RCT using a 2x2 factorial design, participants are allocated to one of four possible combinations: [1] treatment A; [2] treatment B, [3] treatment A and B, or [4] no treatment.

With cross-over trials there may be an interaction between interventions, and this may have an impact on sample size requirements of the study. Interactions are more common when trial interventions share similarities in their mechanisms of action, and result when the effect of one intervention is influenced by another intervention. For example, in the case of a negative interaction, in which the overall effect of individual interventions is reduced when they are provided together, the sample size would need to be increased to still detect a significant difference.

N of 1 Trials

A criticism of the preceding RCT designs is that they may provide good information on treatment outcome for the 'average' patient but are poorly equipped to provide individual-specific information. Randomized trials in individuals are possible, and require limited resources. Such "n of 1" studies are conducted by systematically varying the management of a patient's illness during a series of treatment periods (alternating between experimental and control interventions) in order to confirm the effectiveness of treatment in the individual patient. The number of pairs of interventions often varies from two to seven, but this number is not specified in advance and the clinician and patient can decide to stop when it becomes clear that there are, or are not, important differences between interventions.

Figure: Parallel Trial Design

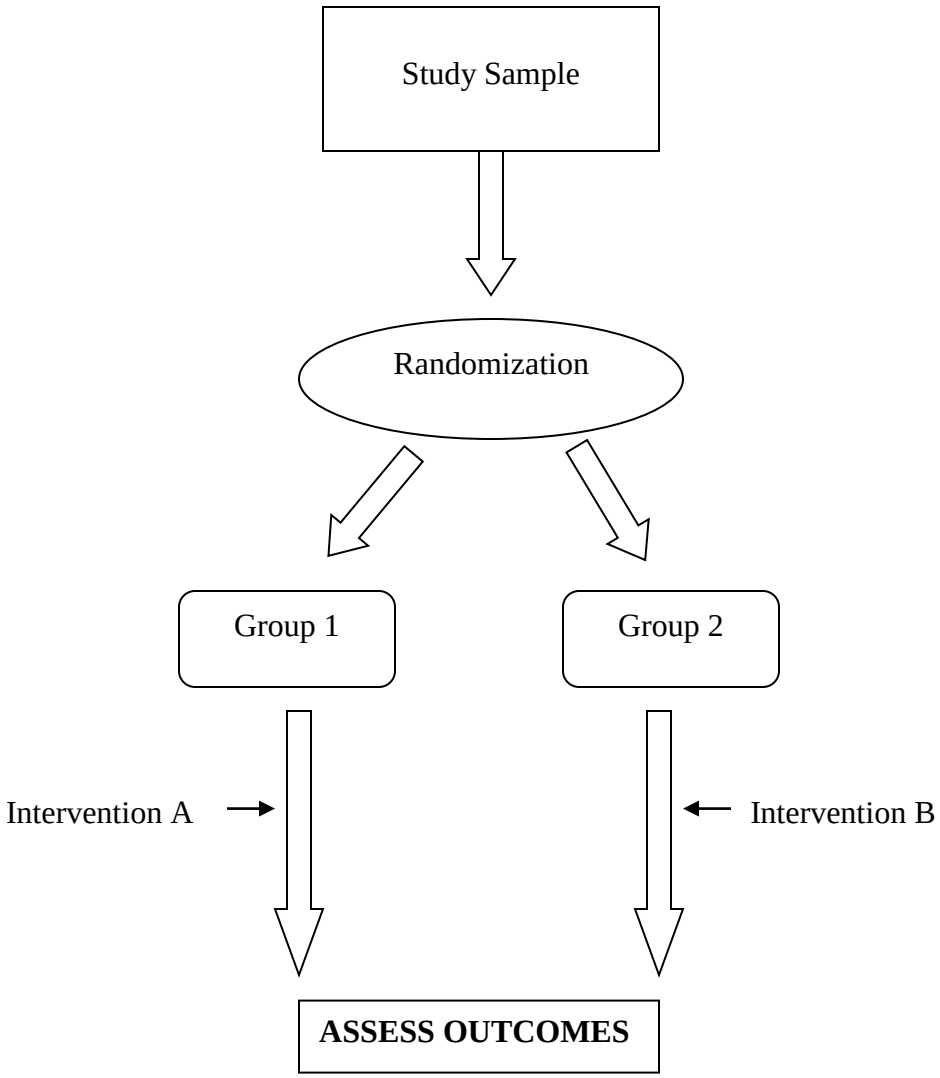


Figure: Cross-over Trial Design

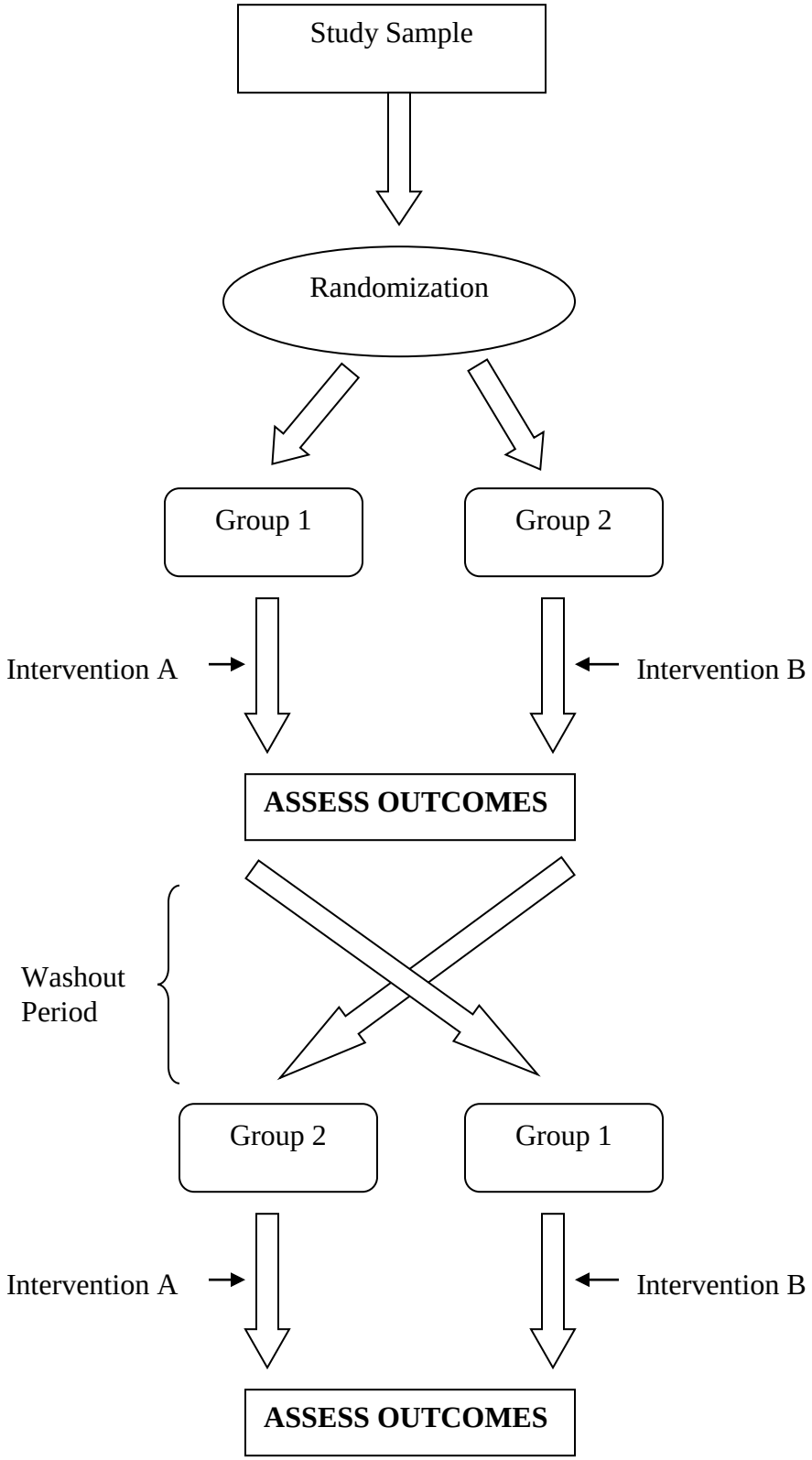


Figure: 2x2 Factorial Trial Design

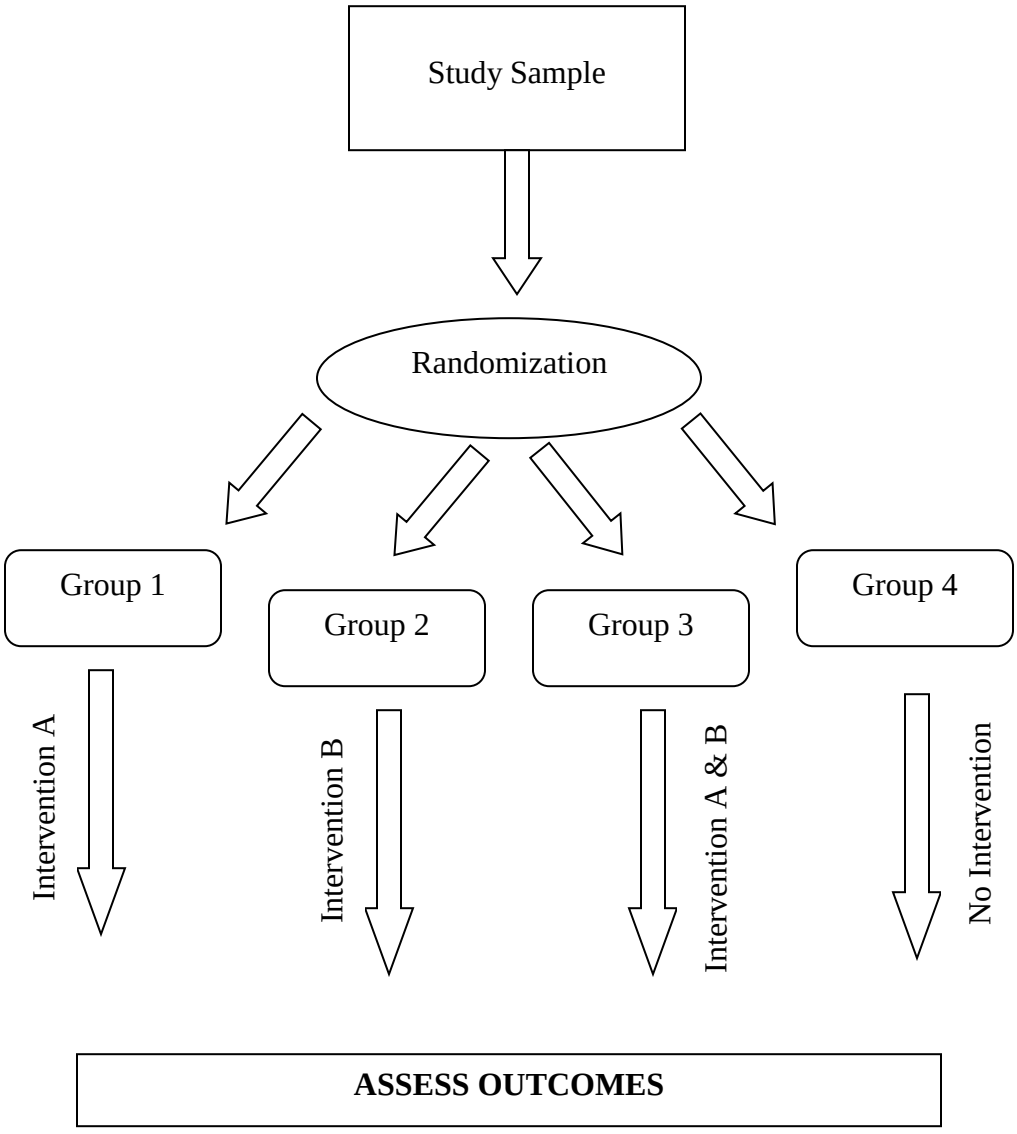
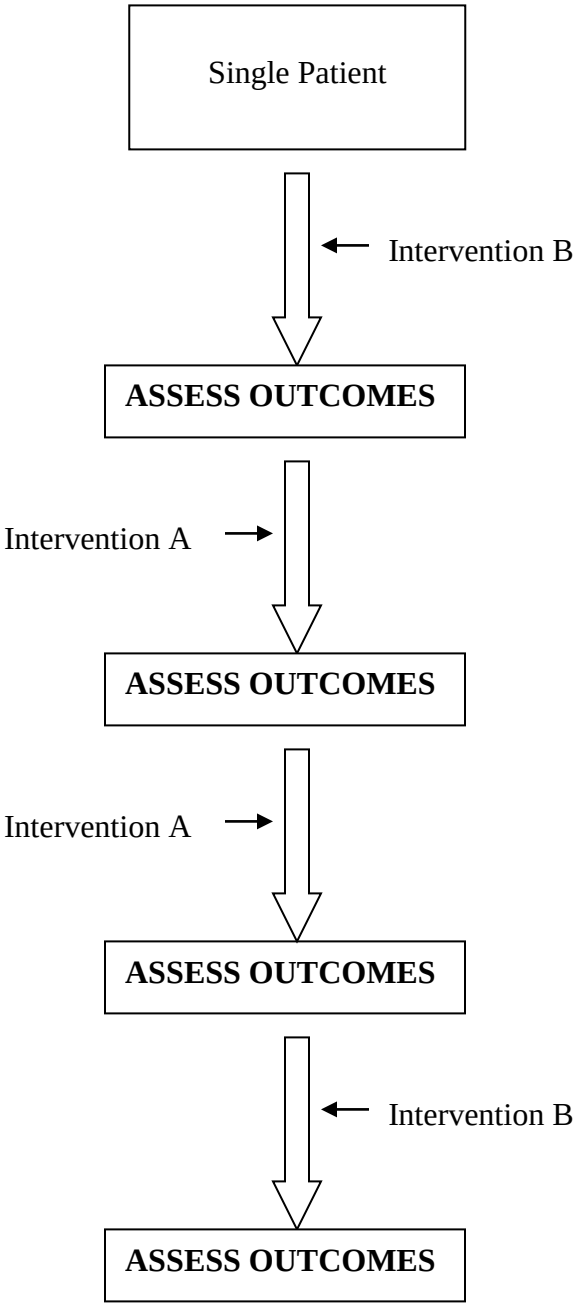


Figure: n-of-1 Trial Design



Observational Studies

Introduction

Assessment of orthopedic treatments depends primarily on studies that analyze data abstracted from medical records or collected prospectively for research purposes. Because treatment in these studies is determined by the surgeon and the patient rather than controlled by the researcher, these studies are referred to as observational. In contrast, protocols in controlled trials specify the details of the treatment including treatment assignment for each patient. The best controlled trials are randomized controlled trials which randomly assign patients to the treatments being compared. This section addresses the main methodological issues in the design and subsequent interpretation of observational studies.

Prospective Cohort Study Design

The term cohort comes from the Roman word for a group of soldiers that marched into battle together. In the cohort study design, the cohort represents a group of people followed over time to see whether an outcome of interest develops. Ideally this group meets a level of certain predetermined criteria representative of a population of interest and is followed with well-defined outcome variables. The Framingham Heart Study is an example of a large cohort study involving residents of a Massachusetts community with identifiable cardiovascular risk factors being followed up for cardiovascular events. Usually this group is matched with a control population selected on the presence or absence of exposure to a factor of interest. The purpose of this type of study is to describe the occurrence of certain outcomes with time and to analyze associations between prognostic factors and those outcomes.

Cohort studies can be prospective in nature meaning that they begin at a specified point and are followed forward in time to evaluate the influence of certain prognostic factors or interventions on the desired outcomes. Examples include prospective cohort studies such as one evaluating refractures in patients initially treated for a fracture. The strengths of a prospective cohort study are the ability for the investigator to study several outcomes with time, and ensure that the data collected are relevant and accurate. The drawbacks are the expense of tracking a large number of subjects and requirement of a long study period. Dropouts increase with time followed and a large number of dropouts can invalidate some cohort study results.

A retrospective cohort or historic cohort involves identifying patients from past records and following this group backward in time from the present to those past records. Retrospective cohort studies have the advantage of being shorter in duration compared with prospective studies but they lack the ability to control the selection of subjects and lack the control over outcome measurements. Cohort studies are observational in nature. Confounding variables may be introduced because random allocation is not used thereby affecting the outcome rather than the factors being examined.

Case-Control Study Design

One type of observational study is the case-control study that starts with the identification of individuals who already have the outcome of interest, cases, and are compared with a suitable control group without the outcome event. The relationship between a particular intervention or prognostic factor and the outcome of interest is examined by comparing the number of individuals with each intervention or prognostic factor in the cases and controls. Case-control studies are described in greater detail later in this report. One example of a case-control study to investigate prognostic factors would be to identify patients with non-healing fractures and a similar group of patients with well-healed fractures to see whether the patients with non-healing fractures were more likely to smoke.

Case-control studies are described as retrospective because they are done looking back in time at information collected about past exposure to possible attributable factors. From this information, an odds ratio can be calculated to describe the odds of a particular factor in individuals with the outcome of interest compared with those without the outcome. The case-control study can be useful in studying rare outcomes, outcomes with multiple potential etiologic factors, or looking at outcomes that take a considerable length of time to develop. Additionally, they can be conducted in a short time with small sample sizes, and for less money than other types of studies. However, because the information usually is collected from patients or their hospital records, data may be inaccurate because of the tendency of those with a disease or bad outcome to ascribe this outcome to exposures (called recall bias) and lack of or biased information form medical records.

Clinical Case Series

Case reports are an uncontrolled, descriptive study design involving an intervention and outcome with a detailed profile of one patient. An early example from the orthopaedic literature of a case report is Birkett's 1869 description of a fracture-dislocation of the hip. Expansion of the individual case report to include multiple patients with an outcome of interest is a case series. In 1981, a famous case series involving five homosexual men in Los Angeles, CA with Pneumocystis carinii between 1980 and 1981 marked the beginning of the AIDS epidemic in the United States. Although descriptive studies are limited in their design to make causal inferences about the relationship between risk factors and an outcome of interest, they are helpful in developing a hypothesis that can be tested using an analytic study design.

Choosing a Study Design

Introduction

In order to plan a clinical study, investigators need to understand the hierarchy of evidence and language of evidence-based medicine to realize how the results of their study may (or may not) be applied to patient care, as previously discussed. A quality clinical study often begins with a quality research proposal and study plan. While there are hundreds of steps required in the development of a study plan, we focus on important considerations that should be addressed in all study plans.

***Step 1**: Asking a Clinically Important Study Question*

The question formulation typically includes a description of the population, intervention, outcomes. The question will also determine the most appropriate study design. For example, one might pose the following: "What is the effect of arthroplasty versus non-internal fixation [*intervention*] on revision rates [*outcome*] in patients aged >65 years with displaced fractures of the femoral neck [*population*]". This question may best be answered with a randomized trial.

***Step 2**: Conducting a Comprehensive Literature Review*

Prior to committing large amounts of time, personnel, and funding to a project, investigators must ensure that their proposed study is novel and advances the current understanding of a problem. A careful and systematic review of the available literature can inform investigators about the current evidence to date. A well-conducted systematic review (and/or meta-analysis) is invaluable since it is unusual for single studies to provide definitive answers to clinical questions. Moreover, a well-conducted quantitative review may resolve discrepancies between studies with conflicting results. We use the term *systematic review* for any summary of the medical literature that attempts to address a focussed clinical question, and meta-analysis as a term for systematic reviews that use quantitative methods (i.e., statistical techniques) to summarize the results guiding principles in the conduct of meta-analyses include a specific health care question, a comprehensive search strategy, the assessment of the reproducibility of study selection, the assessment of study validity, evaluation of heterogeneity (differences in effect across studies), inclusion of all relevant and clinically useful measures of treatment effect.

Because conducting a comprehensive review of the literature comes at the cost of time, effort, and other priorities, surgeons can also seek information from sources that explicitly publish evidence summaries, critically appraised topics and systematic reviews. The practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research. The User's Guide to the Medical Literature that has appeared in the Journal of the American Medical Association and the recent installments of the User's Guide to the Orthopaedic Literature in the Journal of Bone & Joint Surgery provide clinicians with the tools to critically appraise the methodological quality of individual studies and apply the evidence. To provide clinicians with easy access to the best available evidence, several specialized sources include summaries of individual studies, systematic reviews, and evidence-based clinical guidelines. One such example is the Cochrane Database, which is an extensive database of systematic reviews on various topics in musculoskeletal disease. Additionally, the Cochrane Database contains a Controlled Clinical Trial Registry, which provides a comprehensive list of randomized clinical trials in orthopaedics and other subspecialty areas. The Canadian Journal of Surgery, The Journal of Bone and Joint Surgery, and the Journal of Orthopaedic Trauma all provide evidence summaries on a variety of topics.

Table: Potential Information Resources

The Cochrane Library (www. update-software.com)
Bandolier
Best Evidence
University of York/NHS Center for Reviews and Dissemination
Medline
Ovid
HIRU (www.hiru.mcmaster.ca)
Evidence based medicine center at Oxford
Evidence based medicine
ACP journal club

However, even this alternative requires that surgeons possess the skills necessary to understand the available evidence. For instance, to help patients weigh the risks and benefits of a treatment, clinicians must understand the best estimate of the magnitude of the treatment's effect (relative risk) as well as the precision of that estimate (95% confidence interval).

One way to express the impact of one treatment against another is the *relative risk*: the risk of infection among patients on the new treatment, relative to that among controls, For example, let us assume that treatment A results in a 10% infection rate compared with treatment B that causes a 20% infection rate. The relative risk of infection with treatment A is therefore 0.50 (A/B = 0.10 / 0.20 = 0.50). The most commonly reported measure of dichotomous treatment effects is the complement of this relative risk, and is called the *relative risk reduction (RRR)*. It is expressed as a percent: (1 - A/B) x 100% = (1 - 0.50) x 100% = 50%. A RRR of 50% means that treatment A reduced the risk of infection by 50% relative to treatment B; the greater the relative risk reduction, the more effective the therapy.

We usually (though arbitrarily) use the 95% confidence interval. You can consider the 95% confidence interval as defining the range that includes the true relative risk reduction 95% of the time. You'll seldom find the true RRR toward the extremes of this interval, and you'll find the true RRR beyond these extremes only 5% of the time, a property of the confidence interval that relates closely to the conventional level of "statistical significance" of $p < 0.05$.

Step 3: *Refining the Study Question*

Frequently, a comprehensive literature review will identify key reviews or evidence summaries that will assist investigators in planning their study. In some circumstances, the question being posed (see step 1) may have to be re-stated to be most consistent with the current controversies in the research field. For example, a review of the literature to identify evidence evaluating arthroplasty versus internal fixation in patients with displaced femoral neck fractures identifies two meta-analyses. Both studies provide persuasive evidence that arthroplasty reduces the risk of revision surgery by at least 58% (but as much as 87%) compared with internal fixation. However, the effect of arthroplasty versus internal fixation on mortality rates remains uncertain. It would seem appropriate, then, to revise the study question to evaluate mortality rates as the primary outcome of interest instead of revision rates.

***Step 4**: Deciding on the Best Study Methodology (Case Series to Randomized Controlled Trial)*

As previously stated, the question informs the study design. If an investigator wishes to compare the results of two or more treatment strategies, a randomized control trial is the best option; however, if one aims to identify predictors of mortality in patients treated for hip fractures, a case-control study may be designed to compare prognostic variables in patients who died compared with those who lived. If an investigator plans to compare outcomes following operative versus conservative treatment of spinal fractures, a prospective cohort may be best suited. Randomizing patients in this circumstance may be deemed unethical. The strengths and limitations of each study design are presented in the table below.

Key References

American Medical Association: User's Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. In Guyatt GH, Rennie D (eds). Ed 2. Chicago, American Medical Association Press, 2001.

Bernstein J, McGuire K, Freedman KB. Statistical Sampling and Hypothesis Testing in Orthopaedic Research. Clin Orthop. 2003;413:55-62.

Bhandari M, Whang, W, Kuo JC, Devereaux PJ, Sprague S, Tornetta P. The Risk of False-Positive Results in Orthopaedic Surgical Trials. Clin Orthop. 2003;413:63-69.

Bhandari M, Guyatt GH, Swiontkowski MF. User's guide to the orthopaedic literature I: How to use an article about a surgical therapy. Journal of Bone and Joint Surgery- American 2001; 83A: 916-926.

Bhandari M, Guyatt GH, Swiontkowski MF. User's guide to the orthopaedic literature II: How to use an article about prognosis. Journal of Bone and Joint Surgery 2001; 83A: 1555-64.

Bhandari M, Guyatt GH, Montori V, Devereaux PJ, Swiontkowski MF. User's Guide to the Orthopaedic Literature III:How to Use a Systematic Literature Review. J Bone Joint Surg 2002: 84A:1672-82.

Bhandari M, Guyatt GH, Montori V, Swiontkowski MF. User's Guide to the Orthopaedic Literature IV: How to use an article about a diagnostic test. J Bone Joint Surg Am. 2003; 85A:1133-40.

Bhandari M, Morrow F, Kulkarni AV, Tornetta P, 3rd: Meta-analyses in orthopaedic surgery. A systematic review of their methodologies. J Bone Joint Surg 2001; 83A:15-24.

Bhandari M, Schemitsch EH. Randomized trials: a brief history and modern perspective. Tech Orthop 2004; 19:54-56.

Bhandari M, Tornetta P. Issues in the hierarchy of study design, hypothesis testing, and presentation of results. Tech Orthop 2004:19:57-65.

Beaton D, Schemitsch E. Measure of Health-Related Quality of Life and Physical Function. Clin Orthop. 2003 Aug;413:90-105.

Brighton B, Bhandari M, Tornetta III P, Felson DT. Hierarchy of Evidence: From Case Reports to Randomized Controlled Trials. Clin Orthop. 2003;413:19-24.

Busse JW, Bhandari M, Schemitsch EH. Types of randomized trials in surgery. Tech Orthop 2004; 19:77-82.

Devereaux PJ, McKee MD, Yusuf S. Methodologic Issues in Randomized Controlled Trials of Surgical Interventions. Clin Orthop. 2003;413:25-32.

Guyatt GH: Evidence-based medicine: A new approach to teaching the practice of medicine. JAMA 1992;268: 2420-2425.

Hartz A, Marsh JL. Methodologic Issues in Observational Studies. Clin Orthop. 2003;413:33-42.

Jackowski D, Guyatt G. A Guide to Measurement. Clin Orthop. 2003 Aug;413:80-89.

Kocher MS, Henley MB. It is Money That Matters: Decision Analysis and Cost-Effectiveness Analysis. Clin Orthop. 2003;413:106-116.

Montori VM, Swiontkowski MF, Cook DJ. Methodologic Issues in Systematic Reviews and Meta-Analyses. Clin Orthop. 2003;413:43-54.

Sackett DL, Rosenberg WM, Gray JA: Evidence-based medicine: What it is and what it isn't. BMJ 312: 71-72, 1996.

Schunemann H, Bone L. Evidence-Based Orthopaedics: A Primer. Clin Orthop. 2003;413:117-132.

Preparing a Study Protocol

Introduction

In this section we present a guide to planning a clinical study. A high quality clinical study often requires as much time in planning and preparation as it does in its execution. A quality clinical study often begins with a quality research proposal and study plan. While there are hundreds of steps required in the development of a study plan, we focus on important considerations that should be addressed in all study plans.

Step 1: Asking a Clinically Important Study Question

The question formulation typically includes a description of the population, intervention, outcomes. Please refer to the previous section on choosing a study design for more details.

Step 2: Conducting a Comprehensive Literature Review

Prior to committing large amounts of time, personnel and funding to a project, investigators must ensure that their proposed study is novel and advances the current understanding of a problem. A careful and systematic review of the available literature can inform investigators about the current evidence to date. Please refer to the previous section on choosing a study design for more details.

Step 3: Refining the Study Question

Frequently, a comprehensive literature review will identify key reviews or evidence summaries that will assist investigators in planning their study. In some circumstances, the question being posed (see step 1) may have to be re-stated to be most consistent with the current controversies in the research field. Please refer to the previous section on choosing a study design for more details.

Step 4: Deciding on the Best Study Methodology

Please refer to the previous section on choosing a study design for more details.

Step 5: Eligibility Criteria

Investigators should be explicit about the criteria for including patients into their study. A large and comprehensive list of eligibility criteria will limit the generalizability (external validity) of the study results beyond the specific group of included patients. Thus, one strategy to improve the external validity of a randomized trial is to be inclusive in enrolling a diverse group of patients. Alternatively, in a cohort study where the risk of imbalances between patient groups is high, having a comprehensive list of criteria (inclusion and exclusion) may improve the degree to which patient groups are similar.

Step 6: Choosing the Primary Outcome and Secondary Outcomes

The choice of the outcome measure has important implications on the required sample size. We have found that continuous outcome measures (hospital stay, blood loss, functional scores) require smaller samples of patients to achieve adequate study power than dichotomous outcome variables (mortality, infection rates, re-operation rates). We reviewed study power in published studies with 50 patients or less to examine the effect of outcome variable. We found 76 trials with a sample size of 50 patients or less (29 trials-continuous outcomes, 47 trials-dichotomous outcomes). Studies that reported continuous outcomes had significantly higher mean power than those that reported dichotomous variables (Power= 49% vs. 38%, p=0.042). Twice as many trials with continuous outcome variables reached acceptable levels of study power (i.e., >80% power) when compared to trials with dichotomous variables (37% vs. 18.6%, P=0.04).

Step 7: Determining the Sample Size for the Primary Outcome

The power of a study is the probability of concluding a difference between two treatments when one actually exists. Power (1-β) is simply the complement of the Type II error (β). Thus, if we accept a 20% chance of an incorrect study conclusion (β=0.20), we are also accepting that we will come to the correct conclusion 80% of the time. Study power can be used before the start of a clinical trial to assist with sample size determination, or following the completion of study to determine if the negative findings were true, or due to chance.

The power of a statistical test is typically a function of the magnitude of the treatment effect, the designated Type I error rate (α), and the sample size (N). When designing a trial, investigators can decide upon the desired study power (1-β) and calculate the necessary sample to achieve this goal. If investigators are conducting a post hoc power analysis after the completion of the study, they will use the actual sample size obtained to calculate the study's power.

The magnitude of the effect is, for example, the difference between the mean functional score of the surgically treated group and that of the conservatively treated group. To compensate for the variability of the functional scores in each group (variance or standard deviations about the mean scores), the difference can be divided by the standard deviation of the control group. This resultant value is termed the 'effect size'. Interpretation of the 'effect size' is largely a clinical one. It should represent the point at which a surgeon will change his/her practice if the results are true. There are broad guidelines to the interpretation of effect sizes with 0.2 be a small effect, 0.5 to be a moderate effect, and 0.8 to be a large effect.

Given prevalence of Type II errors in orthopaedic trauma trials, investigators should endeavor to pre-plan estimated sample size requirements based upon conventionally accepted standards for study power (80%) and Type I errors (α=0.05). Small pilot studies on a topic of interest or previously reported literature can be helpful in determining the likely treatment effect.

For example, in planning a trial of alternate strategies for the treatment of tibial shaft fractures, an investigator may identify a systematic review of the literature which reports that time to fracture healing with Treatment A is 120 ± 45 days , while time to healing with Treatment B (control group) can be expected to be up to 140 ± 40 days. The expected treatment

difference is 20 days and the effect size is 0.5 (20/40). From Cohen we know that this is a moderate effect and is likely clinically significant if its really true. The anticipated sample size for this continuous outcome measure is determined by the following equation:

$$N = 2\left\{\frac{(Z\alpha + Z\beta)\sigma}{\delta}\right\}^2$$ where $Z\alpha=1.96$, $Z\beta=0.84$, $\sigma=40$, and $\delta=20$

This study will require approximately 63 patients in total to have sufficient power to identify a difference of 20 days between treatments, if it occurs. An investigator may then audit his/her centers last year and decide if enough patients will present to the centre to meet the sample size requirements.

Let's assume that this same investigator chooses nonunion as the primary outcome instead of time to union. Based upon the previous literature, he/she believes that Treatment A will result in a 95% union rate and Treatment B (control group) will result in a 90% union rate. A different sample size calculation for dichotomous variables is presented below:

$$N = \frac{PA\,(100-PA) + PB(100-PB)}{(PB-PA)^2} \times f(\alpha,\beta)$$ where $PA=95$, $PB=90$, and $f(\alpha,\beta)=7.9$

Now, 869 patients are required for the study to identify a 5% difference in nonunion rates between treatments. An investigator may realize that this number is sufficiently large to prohibit him/her from conducting this trial at one centre, and may elect to gain support at multiple sites for this trial. Please refer to the section on sample size calculations for more details.

Step 8: Identifying the Research Collaborators and Assigning Roles

Once the basic research question, study design, and estimated number of patients has been determined, the research team can be organized. Arguably, developing a team may be considered the first priority before moving forward on question development. However, thinking through the available literature, refining the study question and determining a sample size for a study will have considerable impact on the size of the team and the level of expertise required. For instance, an investigator proposing a randomized trial of hemi-arthroplasty versus bipolar arthroplasty in patients with displaced femoral neck fractures on health-related quality of life (SF-36, SFMA functional scoring instruments) may find that only 80 patients are required for adequate study power. If so, his/her institutional volume may be sufficient to conduct the study at a single center. Alternatively, a study of 800 patients may require multiple site investigators.

A successful research project is supported by a motivated, cooperative, and competent research team. The optimal team is greater than the sum of its parts. Each member should provide specific skills to ensure all competencies are maintained. Typically, a team for a large study should have clinical experts, biostatisticians, methodologists/epidemiologists, administrative personnel, and research coordinators/assistants.

Step 9: Writing a Study Proposal

All well designed studies are based upon a clear and comprehensive proposal for the research. The components of a research proposal guideline for the Canadian Institutes of Health Research are presented in the table below. Input from all research team members is critical. Revisions of the proposal are the rule not the exception. In large, collaborative studies a proposal may undergo 20-30 revisions before it is accepted by all team members including the statisticians, methodologist, clinicians, and other research personnel.

Table: Guide for Randomized Trial Protocols (Canadian Institutes of Health Research)

Part 1: THE NEED FOR THE TRIAL
1. What is the problem to be addressed?
2. What are the principal research questions to be addressed?
3. Why is a trial needed now?
4. Give references to any relevant systematic reviews and discuss the need for your trial in light of these reviews
5. How will the results of this trial be used?
6. Describe any risks to the safety of the participants involved in this trial

Part 2: THE PROPOSED TRIAL
1. What is the proposed trial design?
2. What are the proposed trial interventions?
3. What are the proposed practical arrangements for allocating participants to trial groups?
4. What are the proposed methods for protecting against sources of bias?
5. What are the planned inclusion/exclusion criteria?
6. What is the proposed duration of the treatment period?
7. What is the proposed frequency and duration of follow up?
8. What are the proposed outcome measures?
9. How will the outcomes be measured at follow up?
10. Will a health services questions be measured at follow up?
11. What is the proposed sample size and what are the justifications for the assumptions underlying the power calculations?
12. What is the planned recruitment rate?
13. Are there likely to be problems with compliance?
14. What is the likely rate of loss to follow up?
15. What is the proposed type of analyses?
16. What is the proposed frequency of analyses?
17. Are there any planned subgroup analyses?
18. Has a pilot study been carried out using this design?

Part 3: TRIAL MANAGEMENT
1. What are the arrangements for day to day management of the trial?
2. What will be the role of each investigator or collaborator in this trial?
3. Describe any committees such as the steering committee or data monitoring and safety committee (if applicable)

Step 10: Obtaining Ethics Approval

Strict guidelines for the ethical conduct of clinical and experimental research exist. Investigators must be familiar with their local ethics and institutional review processes. Consent forms must be prepared in advance and approval for research must be obtained prior to study start. The process of gaining institutional and ethical approval for a particular study may require from as little as 4 weeks to as much as 8-9 months. Thus, it is imperative that delays in this process are anticipated and accounted for the timeline for the study.

Step 11: Conducting Preliminary or Pilot Studies

The effort to obtain preliminary data can be overstated. Whether an investigator plans to conduct a randomized trial or a prospective cohort study, conducting a pilot study of a proportion of the total sample size (i.e., 10%) can be invaluable to determining the following: 1) the actual ability to screen and recruit patients, 2) investigator and patient compliance to the study protocol, 3) refinement of surgical technique protocols, and 4) ability to achieve complete follow up. Thus, an investigator planning to conduct a study of 200 patients may elect to conduct a pilot study of 20 patients to determine the feasibility of the proposed methodology.

Step 12: Obtaining Research Funding to Conduct a Larger, Definitive Study

The successful conduct of a clinical study (especially a larger prospective cohort study or randomized trial) often requires funding to enable aspects of the patient enrolment, data collection, and data analysis to be performed. Sources of funding can include local or national peer-reviewed granting agencies, industry-sponsored grants, or foundations with specific research agenda. Whatever the source, the pursuit of sufficient funding is a necessary step to ensure optimal rigor and quality in a large clinical study.

Conclusion

The conduct of a well-designed study requires a large time commitment. Often, in a large study, the planning phase will almost require as much time as the conduct phase. As a rule of thumb, one can assume that a study which will take one year to conduct, will likely take one year to plan (literature search, protocol development and revision, obtaining funding). Attention and detail to the study plan will limit problems during the conduct of the study.

GLOSSARY OF TERMS

Absolute Risk Increase: Difference in the absolute risk (percentage or proportion of patients with an outcome) in the exposed vs. the unexposed. Typically used with a harmful exposure.

Absolute Risk Reduction: Difference in the absolute risk (percentage or proportion of patients with an outcome) in the exposed (i.e., to in intervention) (experimental event rate) vs. the unexposed (control event rate). Use restricted to a beneficial exposure or intervention.

Adjusted Analysis An adjusted analysis takes into account differences in prognostic factors between groups that may influence the outcome. For instance, in comparison between an experimental treatment and control groups, if the experimental group is on average older, and therefore at higher risk of an adverse outcome than the control group, the adjusted analysis will show a larger treatment effect than the unadjusted analysis.

Alpha Error: The probability of erroneously concluding there is a difference between two treatments when there is in fact no difference. Typically, investigators decide on the chance of a false positive result they are willing to accept when they plan the sample size for a study.

Baseline Risk: The risk of an adverse outcome in the control group of an experiment. Synonymous with control event rate (CER).

Bayesian Analysis: An analysis that starts with a particular probability of an event (the prior probability) and incorporates new information to generate a revised probability (a posterior probability).

Before-After Trial: Investigation of an intervention in which the investigators compare the status of patients before and after the intervention.

Bias: A systematic tendency to produce an outcome that differs from the underlying truth.

Channeling effect or Channeling bias: The tendency of clinicians to prescribe treatment based on a patient's prognosis. As a result of the behavior, comparisons between treated and untreated patients will yield a biased estimate of treatment effect.

Data completeness bias: Using the information system to log episodes in the treatment group and using a manual system in the non-computer decision support system group can create a data completeness bias.

Detection bias: The tendency to look more carefully for an outcome in one of two groups being compared.

Incorporation bias: When investigators study a diagnostic test that incorporates features of the target outcome.

Interviewer bias: Greater probing by an interviewer in one of two groups being compared.

Publication bias: Publication bias occurs when the publication of research depends on the direction of the study results and whether they are statistically significant.

Recall bias: Recall bias occurs when patients who experience an adverse outcome have a different likelihood of recalling an exposure than the patients who do not have an adverse outcome, independent of the true extent of exposure.

Surveillance bias: Synonymous with detection bias; the tendency to look more carefully for an outcome in one of two groups being compared.

Verification Bias: Results of a diagnostic test influence whether patients are assigned to a treatment group.

Blind (or Blinded or Masked): The participant of interest is unaware of whether patients have been assigned to the experimental or control group. Patients, clinicians, those monitoring outcomes, judicial assessors of outcomes, data analysts, and those writing the paper all can be blinded or masked. To avoid confusion the term masked is preferred in studies in which vision loss of patients is an outcome of interest.

Bootstrap Technique: A statistical technique for estimating parameters such as standard errors and confidence intervals based on resampling from an observed data set with replacement.

Case Reports and Case Series: Descriptions of individual patients. A study reporting on a consecutive collection of patients treated in a similar manner, without a control group. For example, a surgeon might describe the characteristics of an outcome for 100 consecutive patients with cerebral ischemia who received a revascularization procedure.

Case-Control Study: A study designed to determine the association between an exposure and outcome in which patients are sampled by outcome (some patients with the outcome of interest are selected and compared with a group of patients who have not had the outcome), and the investigator examines the proportion of patients with the exposure in the two groups.

Chi-square Test: A statistical test that examines the distribution of categorical outcomes in two groups, the null hypothesis of which is that the underlying distributions are identical.

Clinical Prediction Rules (or Clinical Decision Rules): A clinical prediction rule is generated by initially examining, and ultimately combining, numerous variables to predict the likelihood of a current diagnosis or a future event. Sometimes, if the likelihood is sufficiently high or low, the rule generates a suggested course of action.

Cointerventions: Interventions other than treatment under study that may be differentially applied to experimental and control groups and, therefore potentially bias the results of a study.

Cohort: A group of persons with a common characteristic or set of characteristics. Typically, the group is followed up for a specified period to determine the incidence of a disorder or complications of an established disorder (prognosis).

Cohort Study (or Cohort Analytic Study): Prospective investigation of the factors that might cause a disorder in which a cohort of individuals who do not have evidence of an outcome of interest but who are exposed to the putative cause are compared with a concurrent cohort who also are free of the outcome but not exposed to the putative cause. Both cohorts then are followed up to compare the incidence of the outcome of interest.

Conditional Probabilities: The probability of a particular state, given another state. That is, the probability of A, given B – Probability (A/B).

Confidence Interval (CI): Range of two values within which it is probable that the true value lies for the whole population of patients from whom the study patients were selected.

Confounder: A factor that distorts the true relationship of the study variable of interest by virtue of also being related to the outcome of interest. Confounders are often compared. Randomized studies are unequally distributed among the groups being less likely to have their results distorted by confounders than are observational studies.

Contamination: Contamination occurs when participants in either the experimental or control group receive the intervention intended for the other arm of the study.

Continuous Variables: A variable that can theoretically take any value and in practice can take a large number of values with small differences between them.

Correlation: The magnitude of the relationship between different variables or phenomena.

Correlation Coefficient: A numerical expression of the strength of the relationship between two variables, which can take values from –1.0 to 1.0

Cost Analysis: If two strategies are analyzed but only costs are compared, this comparison would inform only the resource-use half of the decision (the other half being the expected outcomes) and is termed a cost analysis.

Cost Benefit Analysis: A form of economic analysis in which the costs and the consequences (including increases in the length and quality of life) are expressed in monetary terms.

Cost-Effectiveness Analysis: An economic analysis in which the consequences are expressed in natural units. Some examples would include cost per life saved or cost per unit of blood pressure lowered.

Cost Minimization Analysis: An economic analysis conducted in situations where the consequences of the alternatives are identical, and so the only issue is their relative costs.

Cost-Utility Analysis: A type of cost-effectiveness analysis in which the consequences are expressed in terms of life-years adjusted by peoples' preferences. Typically, one considers the incremental cost per incremental gain in quality adjusted life-years.

Cox Regression Model: A regression technique that allows adjustment for known differences in baseline characteristics between experimental and control groups applied to survival data.

Data-dredging: Searching a data set for differences between groups on particular outcomes, or in subgroups of patients, without explicit a priori hypotheses.

Decision Analysis: A systematic approach to decision making under conditions of uncertainty. It involves identifying all available alternatives and estimating the probabilities of potential outcomes associated with each alternative, valuing each outcome, and, on the basis of the probabilities and values, arriving at a quantitative estimate of the relative merit of the alternatives.

Decision under risk: a decision against nature in which a probability distribution on the states of nature is known.

Decision under uncertainty: a decision against nature with no knowledge about the likelihood of the various states of nature.

Decision Tree: Most clinical decision analyses are built as decision trees. Articles about clinical decision analyses usually will include one or more diagrams showing the structure of the decision tree used for the analysis.

Dichotomous Outcomes: "Yes " or "no" outcomes that either happen or do not happen, such as re-operation, infection, and death.

Dichotomous Variable: A variable that can take one of two values, such as male or female, dead or alive, having suffered an infection or not having suffered an infection.

Direct costs: The costs of all resources that can be traced to a particular intervention.

Economic Analysis: A set of formal, quantitative methods used to compare two or more treatments, programs, or strategies with respect to their resource use and their expected outcomes.

Economic Evaluation: Comparative analysis of alternative courses of action in terms of their costs and consequences. The effect size is the difference in outcomes between the intervention and control groups divided by some measure of variability, typically the standard deviation.

Effect Size: The difference in the outcomes between the intervention and control groups divided by some measure of the variability, typically the standard deviation

Efficiency: Achieving the maximal increment in health benefit for a given quantity of resources.

Evidence-Based Medicine: The conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine requires integration of individual clinical expertise and patient preferences with the best available external clinical evidence from systematic research.

Fold back analysis: The process of solving a decision tree by working backward.

Generalizibility: The ability to generalize the findings of a study to a larger group of similar people.

Hawthorne Effect: Human performance that is improved when participants are aware that their behavior is being observed. In a treatment study, the treatment is deemed effective when it actually is ineffective. In a diagnosis study, the patient does not suffer from the target condition, but the test suggests the patient does.

Hazard Ratio: Investigators may compute the relative risk with time, as in a survival analysis, and call it a hazard ratio, the weighted relative risk over the entire study.

Health-Related Quality of Life: Measurements of how people are feeling or the value they place on their health state.

Incidence: Number of new cases of disease occurring during a specified period of time; expressed as a percentage of the number of people at risk.

Heterogeneity: Differences between patients or differences in the results of different studies.

Intention-to-Treat Principle or Intention-to-Treat Analysis: Analyzing patient outcomes based on which group into which they were randomized regardless of whether they actually received the planned intervention. This analysis preserves the power of randomization, thereby maintaining that important unknown factors that influence outcome are likely equally distributed in each comparison group.

Linear Regression: The term used for a regression analysis when the dependent or target variable is a continuous variable and the relationship between the dependent and independent variables is thought to be linear.

Logistic Regression: A term used for a regression analysis in which the dependent or target variable is dichotomous and which uses a model that relies on logarithms.

Meta-Analysis: An overview that incorporates a quantitative strategy for combining the results of several studies into a single pooled or summary estimate.

Multivariable Regression Equation: A type of regression that provides a mathematical model that explains or predicts the dependent or target variable by simultaneously considering all of the independent or predictor variables.

Multivariate Analysis: An analysis that simultaneously considers a number of predictor variables.

Null Hypothesis: In the hypothesis-testing framework, the starting hypothesis the statistical test is designed to consider and, possibly, reject.

Number Needed to Harm: The number of patients who would need to be treated over a specific period of time before one adverse side effect of the treatment will occur. It is the inverse of the absolute risk increase.

Number Needed to Treat: The number of patients who need to be treated over a specific period of time to prevent one bad outcome. When discussing number needed to treat it is important to specify the treatment, its duration, and the bad outcome being prevented. It is the inverse of the absolute risk reduction.

Observational Studies (or Observational Study Design): Studies in which patient or physician preference determines whether a patient receives treatment or control.

Odds: A ratio of probability of occurrence to nonoccurrence of an event.

Odds Ratio: A ratio of the odds of an event in an exposed group to the odds of the same event in a group that is not exposed.

Opportunity cost: The potential benefit given up when the choice of one alternative precludes the selection of a different alternative.

Power: In a comparison of two interventions, the ability to detect a difference between the two experimental if one in fact exists.

Prognostic Factors: Patient or study participant characteristics that confer increased or decreased risk of a positive or adverse outcome.

Prognostic Study: A study that enrolls patients at a point in time and follows them forward to determine the frequency and timing of subsequent events.

P-value: The probability that results as or more extreme than those observed would occur if the null hypothesis were true and the experiment were repeated over and over.

Quality-Adjusted Life-Year: A unit of measure for survival that accounts for the effects of suboptimal health status and the resulting limitations in quality of life. For example, if a patient lives for 10 years and her quality of life is decreased by 50% because of chronic lung disease, her survival would be equivalent to 5 quality-adjusted life years.

Random Allocation: A sample derived by selecting sampling units (e.g., individual patients) such that each unit has an independent and fixed (generally equal) chance of selection. Whether

or not a given unit is selected is determined by chance, for example, by a table of randomly ordered numbers. Allocation of individuals to groups by chance, usually done with the aid of table of random numbers. Not to be confused with systematic allocation (e.g., on even and odd days of the month) or allocation at the convenience or discretion of the investigator.

Randomized Trial: Experiment in which individuals are randomly allocated to receive or not receive an experimental preventative, therapeutic, or diagnostic procedure and then followed to determine the effect of the intervention.

Relative Risk: Ratio of the risk of an event among an exposed population to the risk among the unexposed.

Relative Risk Reduction: An estimate of the proportion of baseline risk that is removed by the therapy, it is calculated by dividing the absolute risk reduction by the absolute risk in the control group.

Reliability: Refers to consistency or reproducibility of data.

Sensitivity: The proportion of people who truly have a designated disorder who are so identified by the test. The test may consist of, or include, clinical observations.

Sensitivity Analysis: Any test of the stability of the conclusions of a health care evaluation over a range of probability estimates, value judgments, and assumptions about the structure of the decisions to be made. This may involve the repeated evaluation of a decision model in which one or more of the parameters of interest are varied.

Specificity: The proportion of people who are truly free of a designated disorder who are so identified by the test. The test may consist of, or include, clinical observations.

Studies or Study Design: The way a drug study is organized or constructed.
(1) Phase I Studies: Studies that investigate a drug's physiologic effect or ensure that it does not manifest unacceptable early toxicity, often conducted in healthy volunteers.
(2) Phase II Studies: Initial studies on patients, which provide preliminary evidence of possible drug effectiveness.
(3) Phase III Studies: Randomized control trials designed to definitively establish the magnitude of drug benefit.
(4) Phase IV Studies or Postmarketing Surveillance Studies: Studies conducted after the effectiveness of a drug has been established and the drug marketed, typically to establish the frequency of unusual toxic effects.

Treatment Effect: The results of comparative clinical studies can be expressed using various treatment effect measures. Examples are absolute risk reduction, relative risk reduction, odds ratio, number needed to treat, and effect size. The appropriateness of using these to express a treatment effect and whether probabilities, means, or medians are used to calculate them depends upon the type of outcome variable used to measure health outcomes. For example, relative risk

reduction is used for dichotomous variables, and effect sizes are normally used for continuous variables.

Utility Measures: Measures that provide a single number that summarizes all of Health-Related Quality of Life are preference- or value-weighted; these have the preferences or values anchored to death and full health and are called utility measures.