Critical Appraisal:
Clinical Trials and Cohort Studies

EBCP Module #12
Outline of Module

- This module will introduce the following topics....
  - Experimental designs
    - Clinical trials
      - Blinding
      - Randomization
      - Intention to treat
      - Analysis: RR, NNT, NNH
      - External validity
      - Surrogate measures
      - Limitations of clinical trials
  - Observational designs
    - Cohort studies
      - Prospective versus retrospective
      - Limitations of cohort studies
Objectives of Module

- Students who complete this module should be able to correctly define and apply the following terms as they pertain to evidence-based clinical practice.
  - Randomization
  - Blinding
  - Intention to treat
  - External validity
  - Surrogate measures & composite endpoints

- Students who complete this module should be able to...
  - Distinguish between an experimental study and an observational study.
  - Distinguish between a retrospective cohort study and a prospective cohort study.
  - Explain why randomization is the best defense against confounding
  - Calculate and interpret a number needed to treat (NNT) and related statistics.
  - Evaluate inclusion & exclusion criteria in the Methods section of a research study to make decisions about generalizability
Experimental Designs
Next Steps

- In the previous module, you learned a basic framework that you can use to critically appraise medical research.

- This framework suggested that you consider the following points, in order, for study appraisal:
  - Bias
    - Selection bias—identify, estimate effect on results
    - Information bias—identify, estimate effect on results
  - Confounding
    - Identify, determine if controlled for, estimate effect on results
  - Chance
    - p-value and confidence interval

- Here, we will use this framework to make sense of two major study types in the medical literature—clinical trials and cohort studies.
What Are Clinical Trials?
Clinical Trials

As you already know, in clinical trials:
- Research participants are divided into two or more groups.
- The “experimental group(s)” receive an experimental drug, device, or other intervention assigned by the researchers.
- The control group receives either placebo or a standard treatment.
- Groups are followed for development of an outcome or for changes in other clinical parameters.

Clinical trials have multiple unique strengths that make them particularly valuable for questions about therapy and prevention.
- Remember that we have previously identified four question domains: therapy, diagnosis, etiology, and prognosis.
- For questions about the benefits of a new therapy, a clinical trial is the optimal study design.
  - However, clinical trials may not be feasible to answer other types of questions—for example, to answer questions about harm or other situations in which investigators cannot assign exposure (e.g., in a prognosis or etiology study, the disease cannot be assigned).
Clinical Trial Examples

- Numerous important clinical trials have affected clinical practice over the history of medicine.

- Some important recent examples include:
  - Comparison of the incidence of cardiovascular disease in post-menopausal women assigned to receive supplemental estrogen and progestin and those assigned to receive placebo (Women’s Health Initiative—WHI, 2002).
  - Comparison of survival among patients with coronary artery atherosclerotic disease treated with percutaneous coronary intervention (PCI) versus medications alone (COURAGE, 2007).
  - Comparison of oxygen saturation targets on mortality and blindness for extremely low birth weight infants on supplemental oxygen (SUPPORT, 2010).
Phases of Clinical Trials

- While there exist many reasons to conduct clinical trials, trials for new interventions are often conducted in a particular sequence as required by governmental organizations such as the Food and Drug Administration (FDA).

- Phase I studies (sometimes called “first-in-man” studies):
  - Follow animal studies that are used to determine lethal and toxic doses
  - Include a small number of participants who are usually healthy
  - Are used to assess the safety and pharmacokinetic profiles of a new intervention in humans

- Phase II studies:
  - Include a small number of participants from the presumed target population (e.g., who have cardiovascular disease for a cardiovascular drug)
  - Are used to determine treatment efficacy in a small group prior to scaling up to a larger study

- Phase III studies:
  - Include large groups of patients assigned to the new intervention or a control—usually a placebo or the current standard of care
  - Provide evidence of drug efficacy and equivalence/superiority to existing care
Conduct of Clinical Trials

Besides studying the effects of new interventions, there exist many questions that can be addressed by clinical trials:

- Do comparable well-known treatments differ in their outcomes?
- How does changing diagnostic screening test frequency affect patient outcomes?
- Does a lower dose of an existing treatment change outcomes?
- Additional questions can be found at [clinicaltrials.gov](http://clinicaltrials.gov), where completed and ongoing clinical trials are registered. Registration is required by many publications.

Regardless of the question, all clinical trials must adhere to strict ethical guidelines ensuring informed consent and protecting participants.

- Locally, institutional review boards (IRB’s) are responsible for ensuring that researchers follow these guidelines
- These ethical guidelines hold not only during recruitment but at all stages of a clinical trial.
  - Rarely, clinical trials may be stopped early if an intervention is shown to cause undue harm.
  - Clinical trials can also be stopped for:
    - *Efficacy*—where experimental treatment is so beneficial that it would be unjust to keep patients on a control treatment
    - *Futility*—where experimental treatment is not beneficial and it would be unjust to keep patients on it instead of standard of care
Appraising Clinical Trials: Internal Validity
Bias in Clinical Trials

- As you know, there are two major categories of bias in medical research:
  - Selection bias
  - Information bias

- To evaluate *selection bias*:
  - As with all studies, you should make sure to read the inclusion and exclusion criteria and understand how the experimental and control groups were chosen.
  - Ask: Are the groups comparable from the start?
  - You should also look to see if there is any difference in the way that participants are lost to follow-up.
  - Ask: Do the groups remain the same throughout the study?
Bias in Clinical Trials

- To evaluate information bias:
  - You should always understand the nature of the intervention and the methods for obtaining outcomes.
  - Additionally, you should consider some key concepts unique to clinical trials that allow for additional reduction of information bias:
    - Placebo control
    - Blinding
    - Randomization

- A placebo prevents participants and/or researchers from differentiating an inactive control group from the treatment group.
  - Classically, a placebo is a “sugar pill” manufactured to look identical to an oral pharmaceutical agent under study.

- Blinding is a related concept. It describes whether participants, researchers, and/or data analysts are aware of group assignments.
  - A placebo control facilitates the process of blinding
  - It is further discussed on the following slide.
Study Blinding

- In a *single-blinded* study:
  - Participants receiving treatments are unaware of group assignment (experimental group vs. control group).

- In a *double-blinded* study:
  - Participants and providers are unaware of group assignment.

- In a *triple-blinded* study:
  - Participants, providers, and data analysts are unaware of group assignment.

Note that blinding is not always possible. For example, it is not practical to blind a surgeon to an experimental procedure.

However, even in surgery, it may be possible to blind participants to an intervention. This is called “sham” surgery and is the equivalent of a surgical placebo.
Confounding and Randomization

- One of the greatest strengths of clinical trials is that they allow for randomization.
  - Randomization is made possible when exposure to the intervention is determined by the researchers.
  - In a **randomized controlled trial (RCT)**, participants are randomly assigned to the treatment group or control group.

- Randomization randomly distributes participant characteristics between groups and thereby greatly reduces the potential for confounding.
  - Remember that confounding requires an outside variable to be associated with the exposure and the outcome.
    - Randomization helps to assure that no potential “confounders” are more common in the treatment or control group, because ALL characteristics (other than the intervention) should end up randomly distributed. This means that different study groups should be similar across a variety of demographics like age, gender, race, socioeconomic status, etc.
Confounding and Randomization

- Randomization allows researchers to assess the effect of the treatment on the outcome after removing the effects of other factors.
  - **Key Concept:** Randomization is the strongest protection against confounding because in theory it distributes both measured and unmeasured (and even un-measurable!) factors equally between groups.

- Although it is the strongest protection against confounding, randomization is not always perfect.
  - It is most useful with large sample sizes. But it is less likely to evenly distribute small samples.
  - Table 1 (or 2) in most RCT’s allows the reader to evaluate the apparent effectiveness of randomization by comparing various characteristics across groups.
  - Additionally, randomization assumes that randomly assigned groups remain randomly assigned at the end of the study.
    - This can be violated when participants in one group are more likely to drop out during the study period—say, due to the side effects of an intervention under study.
In order to preserve the effects of randomization, participants in RCTs should be analyzed the same way that they were randomized at the beginning of the study.

This is called intention to treat (ITT) analysis.

ITT analysis means that participants are analyzed in the treatment group or the placebo group based on their original assignment—and not on whether they complied with the intended intervention.

Failure to use ITT analysis may reintroduce confounding factors into the results by breaking randomization.

In cases where there are major differences in follow-up between groups, you may see per-protocol analysis used—that is, analysis based on whether participants complied with the original intervention.

However, you should remember that these analyses are no longer randomized and so do not protect as well against confounding.

Per-protocol analysis tends to lead to systematic over-estimation of the effect (type I error)

Therefore, intention-to-treat analysis is preferred over per-protocol analysis
Chance and Statistical Analyses

- After assessing a study for both bias and confounding, it is appropriate to consider the role of chance and the results of statistical analyses.

- There are a variety of ways that clinical trial results may be reported:
  - Categorical results are commonly presented in 2x2 tables and using relative risks.
  - In the case of 3+ study groups, methods such as ANOVA may be used.
  - Results for continuous outcomes may be evaluated with t-tests or their non-parametric equivalents.
  - When relevant, time-to-event analyses may be used with hazard ratios to compare group outcomes.

- In all cases, you should consider whether:
  - Appropriate outcome variables are used
  - Statistical tests are appropriate
  - The primary outcome was pre-specified at the beginning of the study
  - If multiple outcomes are assessed, p-values should be adjusted accordingly
  - p-values and confidence intervals are accurately described by the authors’ conclusions
Example RCT Analysis
Example RCT Analysis

- For the following example, we will consider the 10-year follow-up results of a group of 200 patients aged 55-65 with parental history of Alzheimer’s disease.
  - One half of this group was randomly assigned an experimental drug thought to protect against dementia. The other half received a placebo.
  - The study was double-blinded and there was no loss to follow-up.
  - We will use a binary categorical outcome—development of dementia, as determined by clinical criteria—and calculate incidence rates for each group.
### Example RCT Analysis

<table>
<thead>
<tr>
<th></th>
<th>Disease</th>
<th>No Disease</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>30</td>
<td>70</td>
<td>100</td>
</tr>
<tr>
<td>Not Exposed</td>
<td>60</td>
<td>40</td>
<td>100</td>
</tr>
<tr>
<td>Totals</td>
<td>90</td>
<td>110</td>
<td>200</td>
</tr>
</tbody>
</table>

Incidence in exposed group = \(30/100 = 0.3\)
Incidence in non-exposed group = \(60/100 = 0.6\)
RR = \(0.3/0.6 = 0.5\)

Corresponds to a \((1-0.5)\% = 50\%\) relative reduction in risk (that is, the exposed group is 50\% less likely to have the outcome)
The NNT

- In intervention trials, we may look beyond RR to other useful measures of effect.
  - For example, we may ask how many patients need to receive the treatment to prevent a single case of dementia.
    - This statistic is referred to as the *number needed to treat* (NNT)

- Recall that:
  - The incidence in the exposed group was 0.3
  - The incidence in the unexposed group was 0.6
  - The RR—the ratio of incidences—was 0.5

- In addition to *relative risk*—a ratio—we may calculate *absolute risk reduction* (ARR)—a difference obtained by subtracting incidence rates.
  - ARR = 0.6 - 0.3 = 0.3

- The ARR can, in turn, be used to calculate the NNT.
  - NNT = 1/ARR = 1/0.3 = 3.3
  - By tradition, NNT’s are usually rounded up to a “whole person”. Therefore, on average, four patients would need to be treated with the drug in order to prevent one additional case of dementia.
  - Remember that you should always contextualize this result with what you know from the study: Four patients would need to take the drug for 10 years—the period of follow-up, to prevent one case of dementia.
Elaborating on the NNT

- The ARR and NNT are particularly useful tools for discussing evidence with patients.
  - You may find that these statistics reveal information unavailable in relative comparisons like RR and OR.
  - Outcome measures expressed as ratios can be misleading sometimes

- Imagine two studies of preventative interventions that both give the result of RR=0.5.
  - In the first, incidence in one group was 0.3 and in the other was 0.6.
    - RR=0.5
    - ARR=0.3 → NNT=3.3≈4 (usually, the NNT is rounded up to the next person)
  - In the second, incidence in one group was 0.01 and in the other was 0.02.
    - RR=0.5
    - ARR=0.01 → NNT=100

- Information such as the NNT improves our ability to make treatment decisions.
  - It allows for a more informed consideration of costs (e.g., side-effects) and benefits of an intervention when it is discussed with patients.
  - A similar statistic, the number needed to harm, is sometimes calculated the same way as the NNT but for harm (e.g., intervention side-effects).
    - When compared to the NNT, the NNH can aid in decisions about balancing cost and benefit.
Appraising Clinical Trials: External Validity
Clinical Trials: **External Validity**

- The basic question here is “Can I apply this internally valid study’s results to my patient?”
  - In order to answer that question you need to think about three issues:

1. **Were the patients in the study similar to my patient?**
   - A major concern regarding the external validity of clinical trials is that they often include participants different than those to whom the treatment may be usually offered in practice.
   - Check the Methods section for the inclusion and exclusion criteria:
     - Study patients may have fewer chronic illnesses & comorbidities and may not be using other treatments that your patient may be on
     - The age range of study participants may not match your patient, especially in the case of children or the elderly
Clinical Trials: External Validity

2. Were “real” patient-important outcomes considered, or only surrogate (substitute) measures?*

- Patient-important outcomes would include: death, disability, length of hospitalization, needing dialysis, preservation of sight/hearing, etc.

Vs.

- Disease important outcomes: creatinine level, blood sugar, cholesterol, pulse ox reading, blood pressure, EKG changes, etc.
  - Defined by researchers or clinicians; not necessarily important to patients
  - Often quicker, easier, and cheaper to measure than “real” outcomes, so very common to see these in medical research studies
  - “Surrogate Outcomes” are disease important outcomes that seem to correlate well with real clinical endpoints – but this is not a guaranteed relationship! Examples might include troponin levels, HIV viral load, Hemoglobin A1c levels

3. Are the risks/harms/costs worth the benefits?

- These decisions require good communication with your patient – different things may be important to different people. There isn’t just one answer.
- Calculating NNT and NNH can help with balancing risks and benefits – they are much more intuitive than RR or OR or HR
- Remember that issues of feasibility – availability, cost, insurance coverage, etc. – can be critically important to patients

*Also beware of “composite” (combined) outcomes: some researchers will combine multiple related outcomes together (e.g. mortality, MI, hospitalization, and pacemaker placement might be combined as a “primary composite outcome” in a study of a new treatment for heart failure.) Composite endpoints can make it easier for researchers to see an effect with smaller trials – but they can also be very misleading. In applying study results, clinicians need to separate those endpoints when considering the impact of treatment for patients.
A nice discussion of External Validity, surrogate endpoints & use of NNT in the field of emergency medicine, by Dr. Rahul Patwari (10 minutes):

https://www.youtube.com/watch?v=oB60oMIP1hI
Limits to RCTs

Although clinical trials, and especially randomized controlled trials, represent the strongest single-study design for intervention studies, there are many situations where they cannot be used.

Reasons **not** to use an RCT to investigate a clinical question might include:

- **Logistics:** Some exposures cannot be assigned to patients because they are outside of the control of investigators (e.g., many genetic factors or environmental exposures). This applies mainly to Prognosis and Risk/Etiology domain questions.
- **Ethics:** Some exposures cannot be ethically assigned to patients (e.g., carcinogens, infectious agents, unhealthy lifestyle choices, etc). This applies mainly to Risk/Etiology domain questions.
- **Resources:** RCTs require significant financial resources and time. To get an initial answer to an important question quickly, sometimes a shorter and less expensive study design will be chosen for a “pilot study” on a topic.
- **Utility:** Not every question can be answered with an RCT. For example, in evaluating the effectiveness of a new diagnostic test, the results of the new test must be compared to the results of the gold standard test in all study patients, so randomizing patients into different groups doesn’t make sense.

In these circumstances, other clinical designs, including observational studies, may be used.

The benefits and disadvantages of other study designs will be discussed in this and future modules.
Cohort Studies
What are Cohort Studies?
Cohort Studies: The Basics

- As you have already learned: Participants in a cohort study are selected and grouped based on the presence or absence of an exposure of interest. They are then followed for the development of an outcome of interest.
  - Recall that the exposure in a cohort study is not assigned by the investigator.
    - Thus, a cohort study may be used to study exposures that cannot logistically or ethically be assigned by researchers.

- A cohort study may be conducted prospectively or retrospectively.
  - In a prospective cohort study, participants are selected before outcomes have occurred, are sorted into groups based on exposure, and are followed forwards in time until outcomes develop.
    - For example, researchers might select a group of people, divide them into smokers and non-smokers (exposure) and follow them prospectively for 5 years to see which subjects develop a hip fracture (outcome), then analyze the relationship between hip fractures and smoking.
  - In a retrospective cohort study, participants are selected and sorted based on exposure, then outcomes which have already occurred are assessed.
    - For example, researchers might select all patients seen in a primary care clinic 5 years ago and divide them into smokers and non-smokers (exposure). They review their medical records and see who developed a hip fracture during the past 5 years (outcome that has already occurred), then analyze the relationship between hip fractures and smoking.
    - As a general rule, prospective cohort studies are pre-planned, and subjects are enrolled & baseline data collected before any outcomes develop. On the other hand, retrospective cohort studies are conceived after some people already have the outcome, and researchers often end up using pre-existing data that usually was not originally collected for research purposes.

- You can imagine benefits to each of these approaches.
  - Prospective studies are more rigorous because they are designed to look for outcomes before they occur.
  - However, retrospective studies are much faster, less resource intensive, and may put to use existing data collected for other purposes (e.g., medical records, public health databases, insurance company databases, etc.).
Why use Cohort Studies?

- Cohort studies are uniquely equipped to answer questions like:
  - Are there greater complications in surgeries performed by surgeons who operated the night before? (risk/etiology domain)
  - Is abdominal obesity among young men associated with development of cardiovascular disease in middle age? (risk/etiology domain)
  - What is the life expectancy of patients with bipolar disorder compared to the general population? (prognosis domain)

- These examples highlight the necessity of cohort studies as an alternative to RCTs for many questions.
  - It is not possible or appropriate for researchers to assign some exposures (such as bipolar disorder).
  - Cohort studies best capture “real life” exposures (such as operating the night before).
  - They are also well-equipped for long follow-up and studies of prognosis.
Examples of Cohort Studies

- **Well-known long-term cohort studies include:**
  - *The Framingham Heart Study:* Started in 1948 and on-going, with thousands of participants, it includes multiple generations related to the original group based in Framingham, Massachusetts. It was designed to identify major cardiovascular disease risk factors and provided foundational evidence used in clinical practice regarding lipids and blood pressure.
  
  - *British Physicians Study:* Conducted from 1956-2001, it included most physicians in the UK at the time of initiation. It was designed to study the effect of tobacco use on various outcomes and provided the evidentiary basis for linking smoking and lung cancer.
  
  - *Nurses Health Study:* Started in 1976 and on-going, with hundreds of thousands of U.S. nurses enrolled. It was originally designed to investigate the long-term effects of oral contraceptives and was greatly expanded to consider the effects of dietary and lifestyle factors on chronic disease.
Appraisal of Cohort Studies
Appraisal of Cohort Studies

- Appraisal for cohort studies is similar to appraisal for clinical trials.
  - Both follow groups for extended periods and allow us to compare the incidence of an outcome between exposure groups.
  - However, cohort studies are Observational, and clinical trials are Experimental

- The primary difference between clinical trials and cohort studies is assignment of the intervention by the researchers.
  - This difference results because cohort studies are unable to benefit from randomization, blinding, or placebo control.
    - Therefore, clinical trials are better designed to avoid confounding and are the optimal design wherever possible.
    - As previously mentioned, cohort studies have other strengths including their “real world” setting and the ability to study certain exposures that cannot be assigned by researchers.
  - Cohort studies require much more attention to confounding through study design and statistical analysis because they cannot be randomized.

- Some considerations for the appraisal of cohort studies are found on the following slide.
Notes on Appraising Cohort Studies

- **Bias**
  - **Selection bias**
    - Consider who joined the study.
      - If prospective: Did any eligible participants refuse? Why?
      - If retrospective: Who is included in the existing dataset? Does this affect results?
    - Don’t forget about assessing whether loss to follow-up was different in the groups studied.
  - **Information bias**
    - Do the researchers provide an accurate measurement of exposure?
    - Was exposure measurement comparable between groups?
    - Since blinding and placebo control are not possible, could knowledge of exposure status influence exposure measurement?

- **Confounding**
  - Since randomization was not possible, were appropriate methods used to sufficiently control for important potential confounders?

- **Chance**
  - What is the primary outcome? Are the p-value and confidence interval affected by multiple testing?
Clinical Trials versus Cohort Studies

- When compared to clinical trials, the chief limitation of cohort studies is that exposures cannot be assigned by researchers.

- Besides accurate measurement of exposure, clinical trials provide the opportunity for randomization.
  - In cohort studies, the problem of confounding must be addressed using methods such as matching, stratification, and multivariable regression.
    - Unlike randomization, these methods cannot prevent against confounding by unmeasured or unknown factors.

- However, cohort studies permit the study of exposures that cannot be assigned by researchers for logistic or ethical reasons.
  - They also permit the study of questions about “real world” clinical practice and long-term prognosis.
Key Points for Module 12

- In an experimental study such as a clinical trial, investigators assign the exposures or interventions of interest.

- In an observational study, such as a cohort study, investigators observe but do not assign exposures.

- Clinical trials are often randomized and blinded, which helps prevent systematic error from confounding and bias. RCT’s provide the best evidence for questions where exposures, such as treatments, can be assigned.

- Evaluation of external validity, or generalizability, of studies involves consideration of inclusion & exclusion criteria, putting emphasis on patient-important outcomes (rather than surrogate or combined outcomes), and balancing risks and benefit.

- Randomized clinical trials should be analyzed by intention to treat methods; clinicians should be suspicious of “per-protocol” analysis.

- Cohort studies may be conducted prospectively or retrospectively. They provide good evidence for questions where an exposure cannot be assigned, including about harm, real-world practice and long-term prognosis.
Please complete the Module 12 quiz

Module Design Team: Martha Carvour, Matthew Rysavy, Timothy Bahr
Revisions & Updates by T. Hegmann