CHAPTER 1

Introduction to Evidence-Based Medicine

CHAPTER OUTLINE

Approach to Dissecting a Clinical Paper 1
Key Results in a Paper 1
Key Factors That Affect Study Validity 2
Common Statistical Vocabulary 2
Common Study Types 5

Approach to Dissecting a Clinical Paper

While each paper can be dissected at great lengths, as a trainee your primary goal is to understand the main results, its implications, and sources of bias. If you can hold a 1000-foot view of these three points, you will be able to begin incorporating evidence-based medicine into your own practice quite soon. Experts will always opine at length on new clinical studies, and it is impractical for you as a young trainee to always do this. The following sections touch upon the core statistical terminology you will encounter, in what we hope is a truly simple manner.

Key Results in a Paper

While there are many results to a paper that can be interesting, here is a short primer on the key results we are most often interested in.

PRIMARY OUTCOMES

Primary outcomes are the key result(s) that the study is designed to investigate. For example, all-cause mortality is a popular primary outcome. Often, the primary outcome is taken as matter of fact, but care should be given to seek context to the outcomes, especially when it is a composite primary outcome.

COMPOSITE OUTCOMES

Very frequently, you will find that the primary outcome is a composite of several outcomes. For example, a study investigating the impact of a new drug on heart failure outcomes may designate its primary outcome as a combination of worsening heart failure or cardiovascular death. These outcomes with more than one outcome are known as composite outcomes. Combining several outcomes, known as end points, allows for a study to be carried out for shorter durations and with fewer participants. However, as you might imagine, composite outcomes lead to difficulty interpreting results since it is difficult to know what particular outcome affected the composite outcome the most.

For example, if a trial aims to investigate whether a medication provides a mortality benefit for patients with heart failure, a composite endpoint of worsening heart failure OR cardiovascular death that meets statistical significance could be driven mostly by worsening heart failure. Thus, when interpreting studies with composite outcomes, pay attention to which of the individual endpoints accounted for the majority ("drove") the significant outcome.

SECONDARY OUTCOMES

Secondary outcomes are those results of interest that help provide context to the primary outcome(s). For example, a primary composite outcome could be driven by one of the variables, and analyzing the secondary outcomes can help elucidate what the primary outcome really means.

Key Factors That Affect Study Validity

The completion and dissemination of a study do not guarantee that the results are accurate nor applicable to clinical practice. As a young trainee, you need not worry too much about the circumstances that lead to a study being practice changing. Your goal should be to understand the key trials, which are mostly well-validated and widely accepted, having been scrutinized by a host of experts in each field. However, here are some basic concepts to begin thinking about as you dive into the evidence behind why we practice medicine the way we do.

THE CONTROL

A study can be designed in ways that artificially boost the strength of a treatment. For example, we now have many medications that form the cornerstones of guideline-directed medical therapy in heart failure with reduced ejection fraction. If a study were to test a new medication, both the control and experimental arms should be on the standard of care medications. Otherwise, we do not know that the benefits of this new medication would even exist on top of current guideline-directed medical therapies.

Another common area for confusion is the use of placebo in the control arm. In most cases, a new treatment should ideally be compared to the standard of treatment, rather than a sugar pill. This also ensures that participants are treated ethically, as if a treatment currently exists, it would be unethical to withhold this from participants regardless of the need for a control arm.

FUNDING

While many people discredit studies that have industry funding, the reality is that it is very difficult in most cases to conduct large randomized clinical trials (RCTs) without some form of industry funding. For example, most of the statin trials were done with industry funding. Thus, it is academically dishonest to jump to a negative view of a study simply due to the source of funding. Every study should be evaluated solely based on the merits of its methodologies.

Common Statistical Vocabulary

Now let us delve into some of the most common statistical terminology that you will encounter. These are by no means comprehensive or technical descriptions and have been simplified for the purposes of warmly introducing you to these terms. As you become familiar with the following

terms, practice using them in your everyday discussions about various studies in settings such as rounds. This will help bolster your confidence and actively improve your statistical prowess.

STATISTICAL SIGNIFICANCE

Statistical significance, one of the most common phrases you will encounter, is simply whether the result from an experimental group is reliably (reproducibly) different from that of the control group. In other words, a result that is statistically significant can be trusted.

P VALUE

In simple terms, the P value represents the strength of evidence behind a result. The smaller the P value, the stronger the evidence. The most common P value is set at $P > .05$, which means that in 95% of cases, your results will be true.

CONFIDENCE INTERVAL

The confidence interval (CI) is the range (interval) of values that will most likely encompass the true value of any result. The most common use of this you will encounter is when comparing two conditions (e.g., experiment vs. control) to see if they are significantly different from each other. CIs that do not overlap in any values are considered significantly different.

SENSITIVITY

Think of sensitivity intuitively as the ability of a test to correctly identifying a target (think "it's so sensitive..."). In other words, sensitivity describes how adeptly a test can identify a "true positive." Hence, you would favor tests that are more sensitive when trying to find a target.

SPECIFICITY

Conversely, sensitivity is how well a test can identify a "true negative." A high specificity means that you can be sure that a negative result will be correct. Note that as the specificity increases, the sensitivity tends to decrease, and vice versa.

POSITIVE PREDICTIVE VALUE

It is easy to confuse this term with sensitivity. Whereas sensitivity is a *proportion*, positive predictive value (PPV) is a *probability*. PPV provides information on how likely a positive result is to be true. A high PPV means that you can be sure that a positive result will be correct.

NEGATIVE PREDICTIVE VALUE

Negative predictive value (NPV) is similar to specificity, but again represents a *probability* rather than a proportion. NPV provides a sense of how likely a negative result is true. You "predict" the value of a negative result.

NUMBER NEEDED TO TREAT

The number needed to treat (NNT) is simply the number of patients who need to be given a certain intervention for one patient to benefit. The lower the NNT, the better an intervention since more patients benefit with the same number of interventions.

NUMBER NEEDED TO HARM

The number needed to harm (NNH) is the number of patients that need to be treated for one patient to experience a "harm" (e.g., side effect). The higher the NNH, the better an intervention since fewer patients are harmed for the same number of treatments.

HAZARD RATIO

Hazard ratios (HRs) deal with the probability of an event (often death) at one point in time. An HR of 1 means there is no difference between the control and experimental groups. An HR greater than 1 means there is less survival (more death, i.e., events) in the experimental group, and vice versa.

RELATIVE RISK

This is an intuitive concept. Relative risk (RR) reflects the chances of an outcome in one group compared to another. For example, an RR of 2 means that an outcome is twice as likely in the intervention arm.

ABSOLUTE RISK

The absolute risk (AR) is akin to the total risk. This is often a better marker than RR because it does not disguise the value of an intervention. For example, a medication may decrease the RR of a disease by 400% which seems very impressive, but if the total (absolute) risk of a disease is 1 in a million, then decreasing this by 400% does not have a great clinical application since the disease is already so rare.

INCLUSION CRITERIA

These are criteria that qualify people for a study, similar to the qualifying times required for the Boston Marathon.

EXCLUSION CRITERIA

These are criteria that make a person ineligible to participate in a study, similar to how a felony may exclude one from serving as an FBI agent.

INTENTION-TO-TREAT ANALYSIS

Intention to treat (IIT) considers the original treatment assignment, also known as "intention." This may be incongruent with the true results when there is large crossover between treatment groups throughout the duration of a study.

ON-TREATMENT ANALYSIS

On-treatment (OT) analysis is analyzing data with respect to what actual treatment was received, rather than what was the intended treatment. This is often seen as a means of addressing large crossovers between treatment groups.

Common Study Types

Clinical evidence is derived from several different study types and methodologies. Though we commonly look to RCTs as the gold standard study design, other methodologies have their own merits and provide evidence in many situations where RCTs are not possible or impractical. Here is a short primer on the two most common study types you will encounter.

RANDOMIZED CLINICAL TRIALS

The gold standard of clinical study types, the RCT allows for the most control over the study design. For instance, a clear cause-and-effect relationship can be established by subjecting one group to a treatment and another group acting as the control. Furthermore, differences in study arms are negated by randomization. While RCTs hold a lot of power and high regard, be sure to check the methods as there are many ways to intentionally and unintentionally guide results in a particular direction. For example, in the famous COMET trial comparing carvedilol to metoprolol, there was a significant setup favoring carvedilol (e.g., using a maximum dose of carvedilol vs. low dose of metoprolol).

COHORT STUDY

A cohort study is a type of prospective observational study, meaning that there are no specific interventions. Instead, the focus is on a group of people who have a particular characteristic(s). While these studies are not as robust as RCTs, they provide advantages particularly when ethical standards do not allow for RCTs. For example, an RCT could never be conducted of a group of people who are forced to smoke or not smoke cigarettes. In short, a cohort study is a great choice for assessing the impact of risk factors on disease incidence.