Reading the Medical Literature #1
Approach to Reading a Primary Research Article

The ability to efficiently read and evaluate research articles is a critical skill for physicians. The following is a step-by-step approach, appropriated from several sources, to reading a primary research article. It is intended to facilitate an efficient analysis of a research article to allow you to quickly get to the essence of the research and decide if it contains findings relevant to your clinical practice. A systematic approach is recommended to avoid getting confused or distracted from the actual quality of the research. At the end of each section you should review the information you have gathered from the article and decide if the article is conveying useful information that is relevant to your interests. Guidelines for this process are given in the “assessment points” below.

INTRODUCTION
The majority of the introduction can be skimmed to determine if the topic is of interest. Focus on the last paragraph, usually the last sentence, which should state the purpose of the research. This statement should identify the key components of the research: the study population, the independent variables, the main outcome (dependent) variable, and the study design.

Organizational tip: Write each of these key components down so you can look for their full definitions in the methods.

Assessment Points:
1. Does the introduction make a logical case for the importance and relevance of the research question?
2. Can you clearly identify the research question or purpose statement?
3. Are you able to identify the study population, the major independent and dependent variables, and the study design?
   If the answer to any of these questions is “no”, you probably should not bother reading any further.

Brief review of the basic primary research study designs:
Case-control study: A study that identifies subjects with the outcome of interest (these are the cases) and control subjects without the outcome of interest. The researchers look to see if these groups had the exposure of interest, or if they differed in their exposure to variables of interest.
Cross-sectional study: A study that observes a defined population at a single point in time or time interval. Exposure and outcome are determined simultaneously. A common example is single survey studies.
Cohort study: A study that identifies two groups (cohorts) of subjects, one of which received the exposure of interest and one that did not. The researchers follow these groups (cohorts) forward in time to see if they develop the outcome of interest. Cohort studies are by definition prospective. The term “retrospective cohort” refers to the technique of analyzing a large data set that has already been collected. The researcher chooses a starting date within the data set and identifies cohorts that differ in their exposure to the independent variable of interest. The researcher then looks “forward” within the data set to see if the cohorts differed in their development of the outcome of interest (the dependent variable).
Randomized controlled clinical trial: A study in which a group of subjects are randomized into an experimental group and a control group. The researcher exposes the experimental group to the exposure of interest. Both groups are then followed and/or assessed for the outcome of interest (the
main dependent variable).

METHODS
1. Identify the subject population. Look for specific inclusion and exclusion criteria.
2. Find the specific definitions of the independent and dependent variables of interest.
3. Identify the primary outcome (dependent variable) of the study.
4. Diagram and/or outline the study protocol and design.

Brief review of internal and external validity:

**Internal validity:** The extent to which you can believe that there is a causal relationship between the independent and dependent variables. The internal validity is determined by how well the study design dealt with confounding factors – variables other than the independent variable that may influence the outcome (dependent variable). Put another way, internal validity is the extent to which the study minimized systemic error or bias.

**External validity:** The extent to which the study findings can be applied to other populations.

*Organizational tip:* Once you have identified the primary outcome, write down other factors that could potentially affect this outcome other than the main independent variable of the study. For example, if the study involves a prevention program for ACL tears, write down factors that could influence the incidence of ACL tears in a group besides the main independent variable (in this example, the prevention program): gender, age, sport, field conditions, level of play, previous injuries, time spent competing, time spent training, etc. Evaluate how the study deals with these variables so they do not confound the relationship between the prevention program and ACL tears.

**Assessment Points:**
1. Is the study population well defined by the inclusion and exclusion criteria? Evaluate external validity of the study – can you generalize from the study population to your population of interest?
2. Is the primary outcome variable clearly identified? If you cannot clearly identify the primary outcome variable, either the article is poorly designed (or written, but the two often occur together because a poorly designed study is difficult to clearly describe) or the authors are purposely obscuring flaws in their design. Either situation indicates that the internal validity of the study is questionable.
3. Is the major outcome variable the most clinically relevant patient oriented outcome?
4. Look for the potentially confounding variables you previously listed. How are these variables dealt with to minimize systemic bias? Question the internal validity of the study if the variables that you think could be potentially confounding were not measured and/or assessed.
5. Can you confidently diagram/outline the study protocol? It is a bad sign if there is not enough information, or you are unable to understand the description of the study protocol and design. Once again, poorly designed studies are difficult to describe, or more cynically, authors will purposely be vague about flaws in the design.

STATISTICS
The statistics section can be the most difficult to understand, primarily because few readers have a strong enough background in statistics to challenge the author’s approach. However, there is information in the statistic section that can be evaluated without extensive statistical knowledge.
1. Statistics are determined by the study design. The statistics section should reinforce the study design and state clearly what statistical method will be used to answer the primary research question.

2. Look at the power analysis. A power calculation requires the following information that should be listed clearly
   a. Identification of the primary outcome variable.
   b. Estimation of the expected change and standard error of the mean (SEM) in the control group.
   c. The smallest change (difference) that the researchers deem important (clinically significant).

Brief review of statistical hypothesis testing – type I and type II errors:

**Type I error:** Incorrect rejection of the true null hypothesis. In other words, concluding that there is a difference between groups, or a change over time, when in reality, there is no difference (or change). The significance is the measure of the chance of a type I error. A significance of $p=.05$ indicates that there is .05 or 5% chance that the groups are found to be different by chance when in reality, there is no difference.

**Type II error:** Failure to reject the null hypothesis. In other words, concluding that there is no difference between groups when in reality there is a difference. The rate of the type II error is called beta, and related to power ($\text{Power} = 1 - \beta$). A power calculation should be done prior to the experiment to ensure that there are enough subjects to confidently detect a pre-determined difference if a difference actually exists. An “underpowered study” is when no statistical difference is found when there is truly a difference because of too few subjects.

**Assessment Points:**

1. Is a power calculation delineated, including the assumptions that went into the calculation? A lack of a power calculation is a sign that the study was not well thought out. Alternatively, it could be a sign that the study was initially part of a larger study. This unto itself is not a bad thing, but it does indicate that the study was not initially designed to answer the research question.

2. Check and make sure that the primary outcome variable designated in the power analysis is the same as the primary outcome variable in the methods and results. Authors sometimes will change what they designate the primary outcome after they get the results because their initial hypothesis did not work out. The more outcome variables that are looked at, the more likely that one will be “statistically significant” by chance.

**RESULTS**

1. The first table in the results section should be a table comparing groups on demographics and other baseline characteristics. Look to make sure that the groups’ baseline data are similar. In cohort studies, the groups should ideally only differ on the exposure (independent variable) of interest.

2. Make sure all subjects who initially entered into the study are accounted for.

3. The second table or figure in the results section should contain the answer to the primary research question.

4. Look and see how the confounders that you previously identified were dealt with, either in the study design (e.g., randomization, matched controls), subject selection (excluding subjects with characteristics that would confound the result), or statistically (e.g., regression analysis).
5. Subgroup analysis or secondary outcome analysis should be considered “hypothesis generating”.

**Assessment Points:**
1. Internal validity: are major potential confounders dealt with adequately?
2. Are all subjects who initially entered the study accounted for? If not, this can be a source of bias.
3. All experimental designs should use an intent-to-treat analysis with rare exceptions. Not using intent-to-treat analysis introduces significant bias.
4. It is a “red flag” if you cannot easily find the answer to the primary research question. An emphasis on secondary outcomes indicates that the authors are searching for “significant” findings.
5. Statistical significance does not necessarily mean clinical significance. Assess the magnitude of the impact of the intervention/exposure against the cost/risks of implementation. For therapy articles, calculate Absolute Risk Reduction (ARR), Relative Risk Reduction (RRR), and Number Needed to Treat (NNT). For diagnostic articles, calculate (or make note of) the sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio. For articles evaluating harm or etiology, make note of the Relative Risk (for randomized trials or cohort studies) or Odds Ratio (for case-control studies).

**Brief review of calculations to determine if the results are clinically important:**

A. Articles evaluating a therapeutic intervention
   a. Absolute Risk Reduction (ARR) = Control event rate (CER) – Experimental Event Rate (EER)
   b. Relative Risk Reduction (RRR) = ARR/CER
   c. Number Needed to Treat (NNT) = 1/ARR

B. Articles evaluating a diagnostic test
   a. Sensitivity = True Positive (TP) divided by True Positive plus False Negative (FN), TP/[TP+FN], or the proportion of people with a disease (injury) who have a positive test
   b. Specificity = True Negative (TN) divided by True Positive + False Positive (FP), TN/[TN + FP], or the proportion of people without the disease/injury with a negative test.
   c. Likelihood Ratio (LLR) for a positive test result = sensitivity/1-specificity
   d. Likelihood Ratio (LLR) for a negative test result = 1-sensitivity/specificity

**High Specificity** means you can “believe” a positive test is truly positive – the test is good at “ruling things in”

**High Sensitivity** means you can “believe” a negative test is truly negative – the test is good at “ruling things out”

**Likelihood Ratio** is the likelihood that a given test result will occur in a patient with the target disorder compared to the likelihood that the same result will occur in a patient without the disorder.

**DISCUSSION**
The discussion section is the opportunity for the author to comment on the significance the study’s findings. Unfortunately, authors and journals often over-rate the significance of their results. A good discussion should put the study’s findings in context of previous studies. It should also include a discussion of the weaknesses of the study.
There are many excellent resources both in print and on-line that describe in-depth analyses of both primary and secondary research articles. I have found the following 3 references to be particularly practical and useful.

