Introduction to Critical Appraisal

EBCP Module #11
Outline of Module

- This module will introduce the following topics....
  - Critical appraisal of medical literature
    - Internal validity
      - Bias
      - Confounding
      - Chance
    - External validity (generalizability)
    - Practicing and personalizing your approach
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.
  - Bias
  - Confounding
  - Generalizability

- Students who complete this module should be able to...
  - Distinguish between internal validity and external validity.
  - Recognize the major threats to internal validity (bias, confounding, and chance).
  - Develop a logical approach to critical evaluation of medical literature, which can be expanded & personalized over the course of clinical training.
Getting Your Bearings
Getting Your Bearings (review)

- In a previous module, you learned that the best way to start in determining whether an article will be useful to clinical practice is to determine:
  - What clinical question does this article attempt to answer?

- You should then consider:
  - In what context was the study conducted?
    - Among whom? Where? When?
  - What was the “exposure” in the study? And the primary outcome of interest? Were there any secondary outcomes?
    - What kind of variables were used?
  - What was the design of the study?
    - Common study designs medical research are summarized on the following slide.
Study Designs

Are the investigators describing patients or analyzing patients in groups?

Descriptive Studies
- Case studies
- Case series

Analytic Studies
Are the investigators simply observing patients, or are they assigning interventions to patients?

Observational Studies
- Cross-sectional
- Case-control
- Cohort

Experimental Studies
- Clinical trials

Are the results of multiple studies combined?
- Meta-analyses
- Systematic reviews

Note: This diagram does not represent all possible designs!
Next Steps

- In previous modules, we focused primarily on developing a basic knowledge of the language and concepts used to discuss the medical literature.

- In this and subsequent modules you will develop knowledge and skills to appraise this literature.

- Appraisal involves:
  - Determining if the study is any good (i.e. likely to be correct).
  - Determining if the study is of any use to your particular patient.

- In the following slides, we will consider different types of appraisal.
Types of Appraisal
Common Appraisal Methods

- Three common appraisal types that you may encounter on the wards are:
  - Convenient appraisal
  - Cynical appraisal
  - Critical appraisal
Common Appraisal Methods

- **Convenient appraisal** involves:
  - Taking everything at face value
  - Reading only the abstract and relying on its conclusions
  - Reading the entire paper and relying on its conclusions

- **Cynical appraisal** involves:
  - Outright rejection of the paper due to one or two questions or concerns about the study, regardless of their importance
  - Ultimately, taking little or no value from most studies, based sometimes on minor design flaws
Common Appraisal Methods

- **Critical appraisal** involves:
  - Systematically considering positive and negative aspects of a study’s approach
  - Evaluating questions and concerns about the study to determine how these should influence the interpretation of results
  - Considering separately, in sequence:
    1. the merits of the study as a piece of research
    2. (only if it is any good) the study’s utility in the care of specific patients
Critical Appraisal

- Obviously, we hope you engage in *critical appraisal* of the medical literature.

- The goal of this and subsequent modules is to provide you with useful skills to do so.
  - This module will provide you with a framework for reading papers.
  - This framework is not a hard and fast set of rules but provides a logical approach to a complicated process.
    - In this way, it is like methods taught for remembering what to look at in chest x-rays—first look at airway, for example, then bones, etc.

- In the first part of this module, we will consider the evaluation of *internal validity*—that is, a process for analyzing whether or not a study used an appropriate methodology so that we can feel confident in the authors’ conclusions.

- In the second part, we will consider *external validity*—that is, a process for analyzing whether or not study results should be generalized to a particular patient or group. (also called generalizability)
Internal Validity
What is Validity?

- Remember that validity:
  - Is a measure of accuracy
  - Indicates how well a measurement reflects a true or actual measurement (more accurate, more valid)
  - May be reduced by systematic error

To assess the *internal* validity of a study, ask whether the results of the study are likely to be *accurate*, (not whether the results can be applied to patients outside of the study).
Assessing Internal Validity

- Internal validity should always be assessed before external validity.

- Internal validity can be assessed in three steps:
  - Assess for bias
  - Assess for confounding
  - Estimate the role of chance

Common Threats to Internal Validity

1. Bias
2. Confounding
3. Chance
Bias

- Remember that bias is defined as **systematic error** which results in an inaccurate statistical estimate or an inaccurate study result.
  - It may arise from any aspect of a study (e.g., study design, sampling procedures, assessment methods, statistical analyses).
  - It may arise with any study design, although specific designs are more (or less) prone to certain types of bias.

- There are two general categories of bias:
  - Selection bias
  - Information bias

Common Threats to Internal Validity
1. Bias
2. Confounding
3. Chance
Understanding Bias

- You may run across names for many different types of bias, such as:
  - Recall bias
  - Lead-time bias
  - Confirmation bias
  - Hawthorne effect
  - And many more....

- However, generally, these specific types can generally be distilled into two categories:
  - Selection bias involves *who* ends up in the study (participant recruitment & dropout or loss of follow-up)
  - Information bias involves *what* ends up in the study (data gathering & analysis process)

- It is our opinion that the best approach to critical appraisal is to think about these two categories of bias for every study as they encompass almost all other categories of bias.
Direction of Bias

For each category of bias (selection and information bias), you should consider the likely effect of the bias on the study results:

- Bias may lead to:
  - Overestimation of effect (type I error)
    - Moves the estimate further away from the null hypothesis (no difference)
    - Generally, this happens if the source of systematic error makes groups look more different than they really are
  - Underestimation of effect (type II error)
    - Moves the estimate closer to the null hypothesis
    - Generally, this happens if the source of systematic error makes groups look more similar than they really are
Selection Bias

- **Selection bias** is systematic error that involves the inclusion (or loss) of subjects in a study.
  - It should be considered with questions such as:
    - Who was sampled, recruited, or enrolled?
    - Were certain individuals more likely to be included?
    - Were certain individuals more likely to drop out or be lost to follow-up?

- All studies have inclusion and exclusion criteria. Usually these are listed in detail in the Methods section.

- The most important consideration in considering selection bias is determining how the selection of individuals affects study results.
  - Ask yourself: Does the selection of subjects likely result in overestimation or underestimation of the relationship between exposure and outcome?
Selection Bias: Example

- Consider a clinical trial evaluating the effect of a new drug on migraine headaches, where 100 subjects receive treatment and 100 receive placebo.
  - At the end of the study, RR is calculated for the presence of headache in each group.
  - Suppose that 30 subjects in the placebo group and no subjects in the treatment group drop out due to illness (possibly headaches).
  - Because these dropouts are not counted in the calculations at the end of the study, this likely decreases the total number of headaches in the placebo group.
  - Selection bias will then underestimate the effect of treatment in this study.
  - Thought question: What would be the likely impact of disproportionate loss due to drop out from illness (like headaches) or side effects in the treatment group?
Information Bias

- *Information bias* is systematic error involving the data collected for a study.
  - It can be evaluated with questions such as:
    - What was the overall study design – was it observational or experimental?
    - Were groups in the study treated similarly?
    - How were the data for exposure, outcome, and other variables collected?
    - Who or what was the source of the data? Is this a reliable source?
    - Were the data accurate? (Are there reasons for systematic overreporting or underreporting? Was there systematic misclassification?)

- As with selection bias, the information related to information bias can be evaluated from the study description found in the Methods section.

- Likewise, information bias must always be evaluated for its potential effects on the study results.
  - Does the source or method of data collection likely result in *overestimation* or *underestimation* of the relationship between exposure and outcome?
Information Bias: Example

- In a case-control study examining various exposures that may be associated with the development of a rare form of leukemia:
  - The study uses a self-reported questionnaire to assess exposure to a variety of exposures among 100 cases and 200 controls.

- Certain exposures may be systematically over-reported. For instance, participants may be more likely to over-report their fruit/vegetable intake or daily exercise participation.
  - This may result in individuals without the exposure appearing as if they had the exposure. This makes the groups look more similar.
  - Systematic over-reporting may result in an underestimation of the effect of the exposure.

- Certain exposures may be systematically underreported. For instance, concern about others’ perceptions regarding some exposures (e.g., illicit drugs) may lead to underreporting.
  - This may result in individuals with the exposure appearing as if they did not have the exposure (appearing in the non-exposed group). This makes the groups look more similar.
  - Systematic underreporting may result in an underestimation of the effect of the exposure, too.
Evaluating Bias

To recap:
- First, identify the potential for selection and information bias.
- Then, determine whether each results in underestimation or overestimation of the effect under study.
- Finally, determine how this affects the validity of study results.

For both selection and information bias, the direction of a bias matters.
- Not all bias should undermine your view of a study’s results.
  - If you know that the effect in a study is likely to be underestimated (because of a potential source of bias), yet there is still a statistically significant positive finding, you may believe this result because the relationship between exposure and outcome should be even been stronger without the effect underestimation that was introduced by the bias.
  - On the other hand, if the effect in a study may be overestimated, you should be skeptical of a statistically significant finding as you cannot be sure it is true.
Confounding

- After bias, you should evaluate confounding.

- Remember from Module 8:
  Confounding occurs when an outside variable is related in some way* to both the exposure under study and the outcome under study.

- Methods to account for confounding include stratification, randomization, matching, and multivariable statistical analysis.

*the relationship may be association or causation, direct or inverse, and may be a previously known or unknown relationship.
Direction of Bias by Confounding

- As with bias, merely identifying *confounders* is not enough.

- Like bias, confounding may lead to:
  - Overestimation of effect (type I error)
    - Generally, this happens if the error makes groups look more different than they really are.
  - Underestimation of effect (type II error)
    - Generally, this happens if the error makes groups look more similar than they really are.

- The direction in which confounding influences the study effect should affect your understanding of a study’s result, just as we discussed in the section on bias.
Difference between bias & confounding

- With **bias**, the focus is on an artifact created by some part of the research process (assembling subjects, collecting data, analyzing data) that produces a spurious result.
  - Bias can be either conscious or unconscious. In epidemiology, the word bias does not imply, as in common usage, prejudice or deliberate deviation from the truth.

- With **confounding**, one feature of study subjects has not been separated from a second feature (known or unknown), and has thus been *confounded* with it, producing a spurious result.

- Both are sources of **systematic error**. However, *bias creates an association that is not true*, while *confounding describes an association that is true, but potentially misleading*. 
If both:
- Bias and confounding have been assessed, and
- Neither bias nor confounding constitutes a serious threat to the internal validity of a study,

Then the role of chance (random error) in research study results should be considered.

This includes looking at:
- Statistical tests
- Their results (p-values, confidence intervals)
Some considerations regarding the role of chance in study findings:

- How were the data analyzed?
- Is the analysis appropriate?
  - Is the data type properly matched with the analysis type?
    - For example, when multivariate analysis is needed, logistic regression should be used when the outcome variable is categorical.
  - Are the assumptions* of the statistical test considered?
    - For example, certain tests can be used only when the data analyzed is normally distributed. Was this issue taken into consideration when a specific statistical test was selected?

If the analysis was appropriate, consider the p-value or confidence interval to estimate the role of chance in the overall study result.

- How statistically significant is the result?
  - Is the p-value less than alpha? Does the confidence interval include the null effect (e.g., OR=1)?
  - If the study showed no difference, is it possible that sample size or effect size were not large enough to achieve sufficient statistical power?
- How much variation in the estimate is indicated by the confidence interval?

*These assumptions are generally not presented in the modules and will not be tested. This is only for your own information.
Evaluating Internal Validity: Review

- After getting your bearings on a published study (basic study design, the question being asked, and exposure & outcome variables) we suggest the following steps to appraise the study results:

1. Evaluate the potential for **bias** to have skewed the results
   - Selection bias—identify, estimate its effect on results.
   - Information bias—identify, estimate its effect on results.

2. Evaluate the potential for **confounding** to have skewed the results
   - Identify confounders, determine if they were controlled for, and estimate their effect on results.

3. Consider the possible role of **chance** in the results
   - Look at p-values and confidence interval(s)
   - Consider the likelihood of Type I and Type II errors

- Specific criteria for each of these factors varies across study designs – future modules will go into more detail
External Validity
External Validity

- External validity is also known as *generalizability*.

- External validity should be assessed only after internal validity.
  - There is no need to generalize an internally invalid (inaccurate) result!

- To assess external validity, determine how well the study findings fit the particular patient or group of patients with whom you are working.
  - Consider study design, sampling methods, and bias.
  - Consider size of positive treatment effects.
  - Consider size & seriousness of negative treatment effects (i.e. side effects/adverse events)
  - Consider patient preferences, values, resources.

- As you can imagine, this is no small task.
  - However, it is one of the most important steps in evidence-based practice!
  - This crucial step keeps EBCP from becoming an “ivory tower exercise” or “checklist competency.”
  - **Optimal patient care is the primary goal of EBCP.**
Practicing and Personalizing Your Approach
Critical Appraisal and Clinical Practice

- This module represents an introduction to critical appraisal, not a comprehensive review of appraisal methods. Every skill in medicine requires practice and EBCP is no exception to this.

- As you continue to learn about study designs, statistical analyses, and incorporating evidence in your clinical work, practice the concepts presented here and apply them to common clinical problems.

- When you read clinical studies, start to look for bias, confounding, and the role of chance and think about how the studies might have been improved to answer the clinical questions more accurately.
Also remember that reasonable, informed people may disagree about the validity of a study—just as doctors might disagree about the significance of physical exam findings or treatment methods.

This is not to argue that all views are equal, but rather to remind you that you should expect disagreement on occasion.

Regardless of how you appraise studies, the most important endpoint is always the care of your particular patients.
Key Points for Module 11

- Common threats to internal validity include bias, confounding, and chance.

- Common types of bias include selection bias and information bias. Bias may lead to overestimation or underestimation of a study result.

- Confounding is a form of systematic error in which a third variable—somehow related to both the exposure and outcome—confuses the study result. Confounding may also lead to either over- or underestimation of an effect.

- Internal validity should be verified before external validity (generalizability) is assessed.
Please complete the Module 11 quiz

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