kpLibrarians

CRITICAL APPRAISAL

A Toolkit for
Critically Appraising
Evidence-Based Literature

WHAT IS CRITICAL APPRAISAL?

Critical appraisal is the process of evaluating a published study for methodological quality, possible bias, and relevance to your patient (or patient population).

It is an important step in evidence-based practice (EBP) because not all published studies are of equal quality, and using a flawed, invalid, or inapplicable study to make clinical decisions is not EBP!

Critical appraisal begins after evidence has been gathered on a search topic, using search techniques designed for EBP.

Steps in the Critical Appraisal Process

Critical Appraisal seeks to answer three key questions about a research study:

Was an appropriate study design used?

Was the study conducted in a way that minimizes the risk of bias?

Is this study relevant to my patient?

WAS AN APPROPRIATE STUDY DESIGN USED?

Choosing the "best" study type depends on what kind of question is being asked. During critical appraisal, it is important to determine if the researchers have used an appropriate study type for their research question.

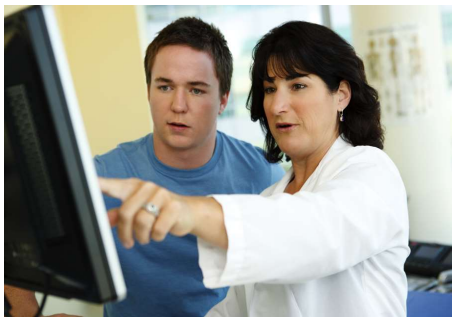
Question Type	Best Study Design
Diagnosis	prospective comparison to the gold standard; cross sectional analytic study
Therapy	randomized controlled trial; review of RCTs
Etiology/Risk/Harm	cohort study; population based case control study
Prognosis	cohort study
Meaning	qualitative study

UNDERSTANDING RESEARCH DESIGNS

Observational Studies

In **observational studies**, researchers just observe what happens to people. They don't intervene in any way other than measuring items of interest.
There are two primary types of observational studies: cohort and case control. In both of these study types, researchers examine two groups of people, divided based on their exposure to something of interest.
In a **cohort study**, researchers follow groups of exposed and unexposed people to see what outcomes develop (or fail to develop). The cohorts are followed forward in time, looking for the development of the outcome.
In **case control studies**, researchers form their groups based on the outcome of interest. They start by assembling a group of cases, (people with the disease or other outcome) and a group of controls (people who could have the outcome, but don't). Then the researchers look back in time to determine exposures.

Randomized Controlled Trials



Randomized controlled trials (RCTs) are experimental studies done to test the effects of an intervention. The intervention can be just about anything the researchers are trying to test: medication, surgery, screening, different delivery methods, etc.
In RCTs, the researchers create two groups of subjects based on a randomization protocol in which people are assigned to either the treatment group or the control group. The groups are then followed to determine how many people in each group develop the outcome of interest.
If done well, RCTs are considered the gold standard of EBM research.

Systematic Reviews

Systematic reviews are literature reviews that use a specific, systematic method to gather data based on already-existing studies. By synthesizing data at a secondary level, systematic reviews can provide a complete picture of the existing evidence on a research question.
Just like with other research designs, systematic reviews must be conducted in a rigorous manner in order to ensure they are providing reliable, evidence-based conclusions. Moreover, the conclusions of a systematic review are only as reliable as the studies on which they are based.

Qualitative Research

Qualitative research involves gathering non-numeric data, such as ideas, opinions, reasons and other non-measurable data. It is often used to gain an understanding of a problem or to explore individuals' experiences.
Qualitative studies typically employ individual interviews, focus groups, or other methods of gathering information about the "whys and hows" of a phenomenon of interest. Although qualitative studies do not involve comparison groups in the same way that cohort, case study, and randomized controlled trials do, they must still be critically appraised to determine the validity of their conclusions.

ARE THE RESULTS THE TRUTH?

The process of **critically appraising** an article helps determine the validity of the results drawn in a research study.

There are four possible explanations for any study result:

- Bias
- Chance
- Confounding
- Truth

If the study result cannot be explained by bias, chance, or confounding, it is most likely a true finding.

Bias in research study refers to a **systematic** error that causes the results to be skewed from the truth. It may not intentional on the part of the researchers, but instead results from a weakness inherent in the study design, or a failure to fully follow best practices in designing the research study. A well-designed study will minimize or mitigate most elements of bias.

There are many types of bias, including selection bias, measurement bias, lead time bias, recall bias, and allocation bias. It is often impossible to completely eliminate bias in a research study, but a close reading of the article (with the help of a critical appraisal checklist) will help detect bias in studies. A listing of biases is available at the Catalogue of Bias.

However, just because a bias exists does not mean the entire study must be immediately disregarded. If a bias cannot be eliminated, part of the appraisal process is to assess the potential impact and take this into account when interpreting study results.

The role of **chance** in a study finding is just what it sounds like -- the possibility that an outcome occurs, regardless of any other factors.

Chance, in this setting, is estimated by determining the possibility that the results would be the same if the study were done again. It is often presented in the article as confidence interval, or **p** value. A **p** value of less than 0.05 indicates that the results likely were not due to chance.

Confounding occurs when there is another variable that is associated with the exposure and outcome being explored, but this element is not part of the cause-and-effect chain being studied. If this variable is unevenly distributed between the groups being studied in a research study, it may confound the results, thereby leading to incorrect conclusions and an invalid result.

It can be difficult to know if a factor is a confounding element, especially because often clinical and scientific knowledge is needed to determine that. As with bias, there are various ways to address (and hopefully eliminate) confounding when creating a study design, but only if the researchers have thought to address it. This is also one of the reasons a well-run randomized controlled trial is so effective at determining causality -- any potential confounding variable will be distributed evenly between the study groups.

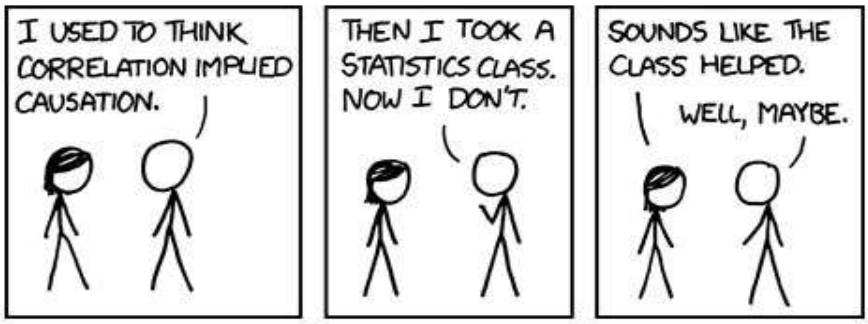
Confounding can also be thought of as an alternative explanation for a study's result.

Correlation vs. Causation

A related issue to correlation and confounding is causation. You may have heard this issue phrased as, "correlation does not equal causation."

Just because two variables are correlated--appear to move together or appear at around the same time--does not automatically mean that the change in one variable **causes** the change in the other variable. They could both be related to a third variable, or appear developmentally around the same time.

There are several criteria to consider when evaluating if the results uncovered by a study are likely to be indicative of a causal relationship. It is not essential that all of these criteria be met, but the more that are satisfied, the greater the likelihood of causality.



WAS THE STUDY CONDUCTED IN A WAY THAT MINIMIZES BIAS?

Bias in research study refers to a **systematic** error that causes the results to be skewed from the truth. It may not intentional on the part of the researchers, but instead results from a weakness inherent in the study design, or a failure to fully follow best practices in designing the research study. A well-designed study will minimize or mitigate most elements of bias.

There are many types of bias, including selection bias, measurement bias, lead time bias, recall bias, and allocation bias. It is often impossible to completely eliminate bias in a research study, but a close reading of the article will help detect the likelihood of bias in a study.

This is where the use of critical appraisal tools and checklist comes in. These tools, along with an understanding of study designs, terminology, and potential flaws, can help detect likely bias in a study.

However, just because a bias exists does not mean the entire study must be immediately disregarded. If a bias cannot be eliminated, part of the appraisal process is to assess the potential impact and take this into account when interpreting study results.

APPRAISAL TOOLS & CHECKLISTS

There are a wide **range of checklists** to help you critically appraise an article.
Be sure to choose a checklist appropriate for the type of study you are evaluating.

- AMSTAR Checklist for Systematic Reviews
  http://amstar.ca/Amstar_Checklist.php
- Cardiff University Critical Appraisal (SURE Checklists)
  https://www.cardiff.ac.uk/specialist-unit-for-review-evidence/resources/critical-appraisal-checklists
- CASP Checklists
  https://casp-uk.net/casp-tools-checklists/
- CEBM Critical Appraisal Tools
  https://www.cebm.net/2014/06/critical-appraisal/
- GRADE Recommendations
  http://www.gradeworkinggroup.org/
- Joanna Briggs Institute Critical Appraisal Tools
  https://joannabriggs.org/critical_appraisal_tools
- Newcastle-Ottawa Scale for assessing non-randomized studies
  http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
- ROBIS: Risk of bias in systematic reviews
  http://xn--robistool-z79d.info/
- Scottish Intercollegiate Guidelines Network checklists
  http://www.sign.ac.uk/checklists-and-notes.html

USERS' GUIDES TO THE MEDICAL LITERATURE

This article series (from the Evidence-Based Medicine Working group and originally published in JAMA) helps the user walk through the critical analysis of articles about therapy, prevention, harm and prognosis, as well as articles on diagnostic tests, clinical decision analysis tools, and other types of studies. Highly recommended reading for understanding studies based on the outcome of interest.

https://www.ncbi.nlm.nih.gov/pubmed/?term=users+guides+to+the+medical+literature+jama

Selected Critical Appraisal Bibliography

Books:

Greenhalgh, Trisha. *How to read a paper: the basics of evidence-based medicine.* 4th ed. Wiley-Blackwell/BMJ Books, 2010.

Guyatt, Gordon, et al. *Users' guides to the medical literature: a manual for evidence-based clinical practice.* 2th ed. McGraw-Hill, 2015.

LoBiondo-Wood, Geri and Judith Haber. *Nursing research : methods and critical appraisal for evidence-based practice.* 9th ed. Elsevier, 2018

Moore, R.A. *Bandolier's little book of making sense of the medical evidence.* Oxford University Press, 2006.

Polit, Denise. *Essentials of nursing research: appraising evidence for nursing practice.* 9th ed. Lippincott Williams & Wilkins, 2018.

Prasad, Kameshwar. *Fundamentals of evidence-based medicine.* 2nd ed. Springer, 2013.

Rebar, Cherie R. *Understanding research for evidence-based practice.* 4th ed. Wolters Kluwer/Lippincott Williams & Wilkins Health, 2015.

Riegelman, Richard K. *Studying a study and testing the test: how to read the medical evidence.* 5th ed. Lippincott Williams & Wilkins, 2005.

Schmidt, Nola A. and Janet M. Brown. *Evidence-based practice for nurses: appraisal and application of research.* 4th ed. Jones & Bartlett Learning, 2019.

Web Sites:

Art of reading a journal article. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3687192/

Catalog of Bias https://catalogofbias.org/biases/

Cochrane community glossary https://community.cochrane.org/glossary

Critical appraisal for medical and health sciences https://www.escholar.manchester.ac.uk/learning-objects/mle/critical-appraisal/story_html5.html

How to read a paper https://www.bmj.com/about-bmj/resources-readers/publications/how-read-paper

NICE algorithm for classifying study designs https://www.nice.org.uk/process/pmg20/resources/appendix-h-pdf-2549710190

Understanding health research: correlation and causation https://www.understandinghealthresearch.org/useful-information/correlation-and-causation-15

Users' guides to the medical literature https://www.ncbi.nlm.nih.gov/pubmed/?term=users+guides+to+the+medical+literature+jama

What is critical appraisal? http://www.bandolier.org.uk/painres/download/whatis/What_is_critical_appraisal.pdf

Glossary of Evidence Terms

Absolute Risk Reduction (ARR): The difference between the percent of people in the control group experiencing a particular outcome and the percent of people in the experimental group experiencing the same outcome. ARR is better able to discriminate between large and small treatment effects than Relative Risk Reduction (RRR).

Bias: Any tendency to influence the results of a study or its interpretation other than the experimental intervention.

Blinding: A technique used in research to eliminate bias by hiding the intervention from the patient, the clinician, and/or the researcher interpreting the results.

Case Control Study: A study which involves identifying patients who have the outcome of interest (cases) and patients without the same outcome (controls) and looking back to see if they had the exposure of interest.

Case Series: A collection of reports on the treatment of individual patients. No control group is involved.

Cohort Study: A study in which patients who presently have a certain condition and/or receive a particular treatment are followed over time and compared with another group who are not affected by the condition under investigation.

Confidence Interval (CI): Quantifies the uncertainty in measurement. It is usually reported as 95% CI which is the range of values within which we can be 95% sure that the true value for the whole population lies. For example, for a NNT of 10 with a 95% CI of 5 to 15, we would have 95% confidence that the true NNT value lies between 5 and 15.

Confounding Variable: A variable which is not the one you are interested in but which may affect the results of the study.

Critically Appraised Topic: A short summary of an article from the literature, created to answer a specific clinical question.

Cross Sectional Study: Study in which a defined population is observed at a single point in time or time interval. Exposure and outcome are determined simultaneously.

Crossover Study: Study in which two or more experimental therapies are administrated one after another in a specified or random order to the same group of patients.

Decision Analysis: The application of explicit, quantitative methods that quantify prognoses, treatment effects, and patient values in order to analyze a decision under conditions of uncertainty.

Event Rate: The proportion of patients in a group in whom the event is observed. If, out of 100 patients, the event is observed in 27, the event rate is 0.27. Control event rate (CER), experimental event rate (EER) and patient expected event rate (PEER) are often calculated as well.

Heterogeneity of Results: Greater variation between the study results in a systematic review than would be expected to occur by chance alone.

Incidence: The proportion of new cases of the target disorder in the population at risk during a specific time interval.

Inception Cohort: A group of patients who are assembled near the onset of the target disorder.

Intention-to-Treat Analysis: A method of analyzing randomization trials in which all patients randomly assigned to one of the treatments are analyzed together, regardless of whether or not they completed or received that treatment, in order to preserve randomization.

Likelihood Ratio: The likelihood that a given test result would be expected in a patient with the target disorder compared with the likelihood that this same result would be expected in a patient without the target disorder.

Magnitude of Benefit: The size of the benefit of the therapeutic intervention being evaluated. The magnitude of benefit is generally defined in terms of the standard deviation of the outcome measure: Large if SD › or = 1; Medium if SD 0.5 to 0.9 orSmall if SD 0.2 to 0.4

Meta-Analysis: A systematic review that uses quantitative methods to synthesize and summarize the results.

N-of-1 Trial: Trial in which the patient undergoes pairs of treatment periods organized so that one period involves the use of the experimental treatment and the other involves the use of an alternate of placebo therapy. The patient and physician are both blinded if possible. Treatment periods are replicated until the clinician and patient are convinced that the treatments are definitely different or definitely not different.

Negative Predictive Value: Proportion of people with a negative test who are free of the target disorder.

Number Needed to Treat: The number of patients who would need to use the therapy under investigation for the period of time described in the study in order for one person to experience the specified benefit. It is calculated by dividing the absolute risk reduction into 1.

Odds Ratio: The ratio of the odds of having the target disorder in the experimental group relative to the odds of having the target disorder in the control group.

***p* Value:** The probability that a particular result would have happened by chance.

Positive Predictive Value: Proportion of people with a positive test who have the target disorder.

Prevalence: The baseline risk of a disorder in the population of interest.

Randomization: Method analogous to tossing a coin to assign patients to treatment groups.

Randomized Controlled Clinical Trial: A true experiment characterized by manipulation of the principal independent variable, random assignment of individual subjects to the treatment and control groups, and the same measurement for both groups.

Risk Ratio (also called Relative Risk): The ratio of risk in the treated group (EER) to the risk in the control group (CER).

Systematic Review: Formal review of a focused clinical question based on a comprehensive search strategy with explicit inclusion criteria and a structured critical appraisal.

Validity: The extent to which a variable or intervention measures what it is supposed to measure or accomplishes what it is supposed to accomplish. The **internal validity** of a study refers to the integrity of the experimental design. The **external validity** refers to the appropriateness by which its results can be applied to non-study patients or populations.