

This shortcut sheet was developed by Research Advocacy Network to assist advocates in understanding Levels of Evidence and how these concepts apply to clinical practice.

Introduction

The concept of levels of evidence is most closely associated with two areas, evidence-based medicine and the development of practice guidelines. The goal of evidence-based medicine is for all doctors to make decisions about the care of individual patients based on "conscientious, explicit, and judicious use of current best evidence."¹

Practice guidelines have been developed to assist physicians in making decisions based on the best available evidence. Two organizations that offer treatment guidelines are:

1. National Comprehensive Cancer Network (NCCN)
<http://www.nccn.org/>
2. American Society of Clinical Oncology (ASCO)
<http://www.asco.org/ac/1,1003,12-002009,00.asp>

To evaluate the strength of different types of clinical evidence, studies or clinical trials are ranked from those with the least amount of bias to those with the potential for the greatest amount of bias. There are many definitions of levels of evidence and when evaluating guidelines reviewing the categories is essential.

Hierarchy of Evidence

In general the hierarchy of evidence is:

1. meta-analysis (quantitative systematic review) using comprehensive search strategies
2. high-quality randomized controlled trial (RCT)
 - a. double blind
 - b. single blind
 - c. non-blinded
3. cohort study
4. case control studies
5. case series and case reports

**For an explanation of the above, please refer to the SUNY Tutorial [Guide to Research Methods](#) included as an appendix in this packet of information.

The hierarchy of evidence will be similar but not identical depending on the question being asked, treatment, prevention, prognosis, diagnosis, etc. Examples of the application of the hierarchy to different situations can be found at:

¹ Evidence-based medicine, Wikipedia: The Free Encyclopedia http://en.wikipedia.org/wiki/Evidence-based_medicine

- ◆ NCI Levels of Evidence for Cancer Screening and Prevention Studies
<http://cancertrials.nci.nih.gov/cancertopics/pdq/screening/levels-of-evidence/healthprofessional/allpages/print>
- ◆ NCI Levels of Evidence for Adult Cancer Treatment Studies
<http://cancertrials.nci.nih.gov/templates/doc.aspx?viewid=2b9ac8c6-7202-4728-9dd0-77ca57170044§ionid=30&version=1&print=1>
- ◆ NCI Levels of Evidence for Human Studies of Cancer Complementary and Alternative Medicine
<http://cancertrials.nci.nih.gov/cancertopics/pdq/levels-evidence-cam/HealthProfessional/page4>

In developing practice guidelines organizations often use panels of experts to analyze the current research and make recommendations. The recommendations are usually accompanied by information about the consensus reached (or not reached) by the panel, called categories of consensus.

The NCCN describes its process this way: *"...the NCCN panels use a consensus process based on expert knowledge and analysis of the most pertinent data. Many of the panels incorporate formal presentations of data for specific controversial issues into the guideline meetings. Because the panels are composed of tumor specialists from each of the member institutions, panel members have in-depth knowledge of the literature and awareness of, if not actual participation in, the trials that provide the evidence for the formulations."*²

NCCN Categories of Consensus

Category 1	There is uniform NCCN consensus, based on high-level evidence, that the recommendation is appropriate.
Category 2A	There is uniform NCCN consensus, based on lower level evidence including clinical experience, that the recommendation is appropriate.
Category 2B	There is non-uniform NCCN consensus (but no major disagreement), based on lower-level evidence including clinical experience, that the recommendation is appropriate.
Category 3	There is major NCCN disagreement that the recommendation is appropriate.

http://www.nccn.org/professionals/physician_gls/PDF/breast.pdf All recommendations are category 2A unless otherwise noted.

² Clinical Practice Guidelines http://www.nccn.org/professionals/physician_gls/about.asp

ASCO Levels and Grades of Evidence³

Level	Type of Evidence
I	Evidence is obtained from meta-analysis of multiple, well-designed, controlled studies. Randomized trials have with low false-positive and low false-negative errors (high power)
II	Evidence is obtained from at least one well-designed experimental study. Randomized trials have high false-positive and/or -negative errors (low power).
III	Evidence is obtained from well-designed, quasi-experimental studies such as nonrandomized, controlled, single-group, pre-post, cohort, time, or matched case-control series.
IV	Evidence is from well-designed, nonexperimental studies, such as comparative and correlational descriptive and case studies.
V	Evidence is from case reports and clinical examples.
Grade	Grade for Recommendation
A	There is evidence of type I or consistent findings from multiple studies of types II, III, and IV.
B	There is evidence of types II, III, and IV, and findings are generally consistent.
C	There is evidence of types II, III, and IV, but findings are inconsistent.
D	There is little or no systematic empirical evidence.

³ <http://www.jco.org/cgi/content/full/17/9/2971/TBL22971>

The following tutorial is reproduced by permission granted to Research Advocacy Network. This tutorial may also be accessed online at: <http://library.downstate.edu/EBM2/contents.htm>

Suny Downstate Medical Center Evidence Based Medicine Tutorial

Guide To Research Methods

<http://library.downstate.edu/EBM2/contents.htm>

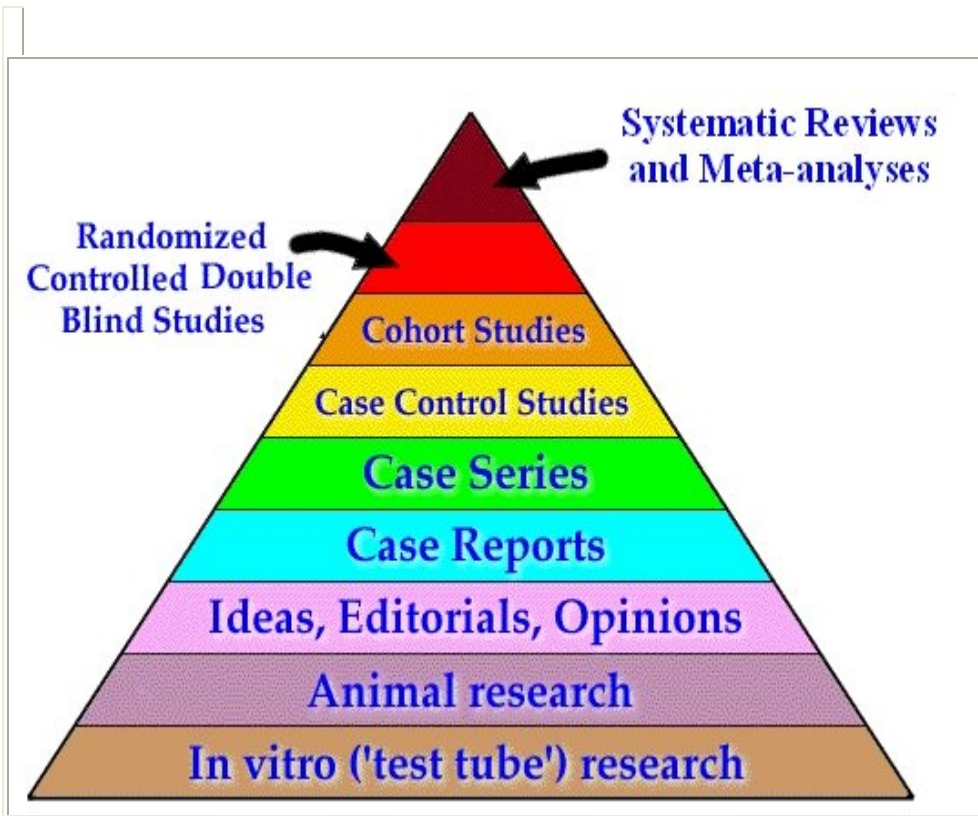
**Copyright © SUNY (State University of New York), 2004
Reproduced with permission**

This tutorial is based on original course material prepared by Betty Wagoner, Reference Librarian and Lead Content Specialist, and Martin Mellish, Co-author and Lead HTML Programmer.

Revisions to this original course material include additions by Anita Andrusek, Mary Doherty, Ki-tai Mok and Andrea Markinson.

This revision includes information and examples presented in lectures by Dr. Eleanor Wallace, Dr. Eugene Dinkevich, Dr. Richard Sinert, Andrea Markinson, and Christopher Stewart.

The Evidence Pyramid

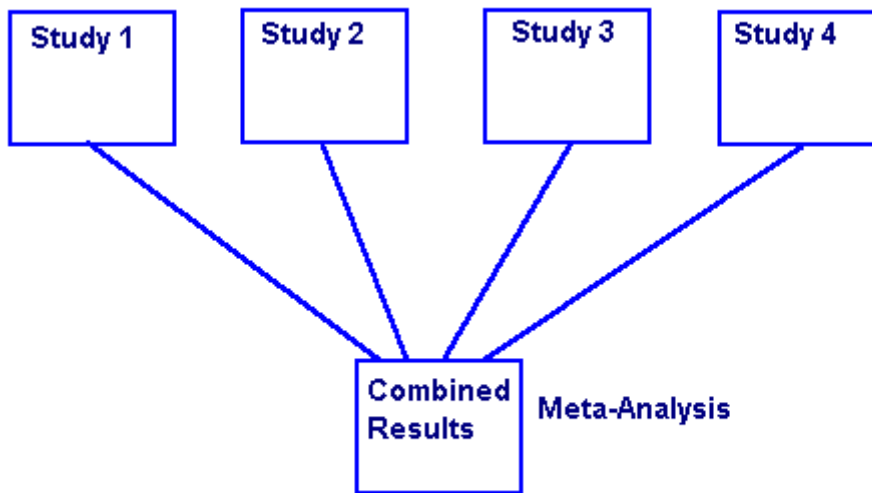


MEDLINE and the other online medical literature databases try to be as comprehensive as possible in their coverage. As a result, indexed material may have little direct application to present-day medical practice.

The different types of material indexed in MEDLINE are labeled in the pyramid diagram, with the least clinically relevant at the bottom and the most clinically relevant at the top. The four layers above case reports and case series represent actual clinical research; the layers below are least clinically relevant and can be useful as background resources.

The next few pages provide basic definitions and examples of clinical research designs to help the medical student or new clinician understand how the design of a research study may affect whether or not to accept its findings in caring for a patient.

Systematic Reviews and Meta-Analyses



Important medical questions are typically studied more than once, often by different research teams in different locations.

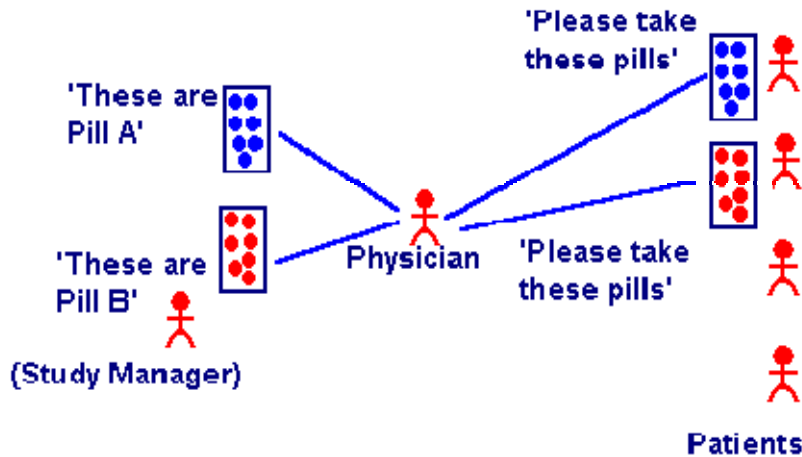
A systematic review is a comprehensive survey of a topic in which all of the primary studies of the highest level of evidence have been systematically identified, appraised and then summarized according to an explicit and reproducible methodology.

A meta-analysis is a survey in which the results of all of the included studies are similar enough statistically that the results are combined and analyzed as if they were one study. In general a good systematic review or meta-analysis will be a better guide to practice than an individual article.

Pitfalls specific to meta-analysis include:

1. It's rare that the results of the different studies precisely agree, and often the number of patients in a single study is not large enough to come up with a decisive conclusion.
2. If the authors are interested in supporting a particular conclusion, they can include studies that support that conclusion and omit studies that do not. Do the authors explain in their paper exactly on what basis they included studies, and do their reasons make sense?
3. Studies that show some kind of positive effect tend to be published more often than those that do not. This means that if the authors include only published studies, several weak positive studies may seem to add up to a strong positive result. Do weak negative studies exist? This effect is known as **Publication bias**.

The Double Blind Method



A double blind study is one in which neither the patient nor the physician knows whether the patient is receiving the treatment of interest or the control treatment.

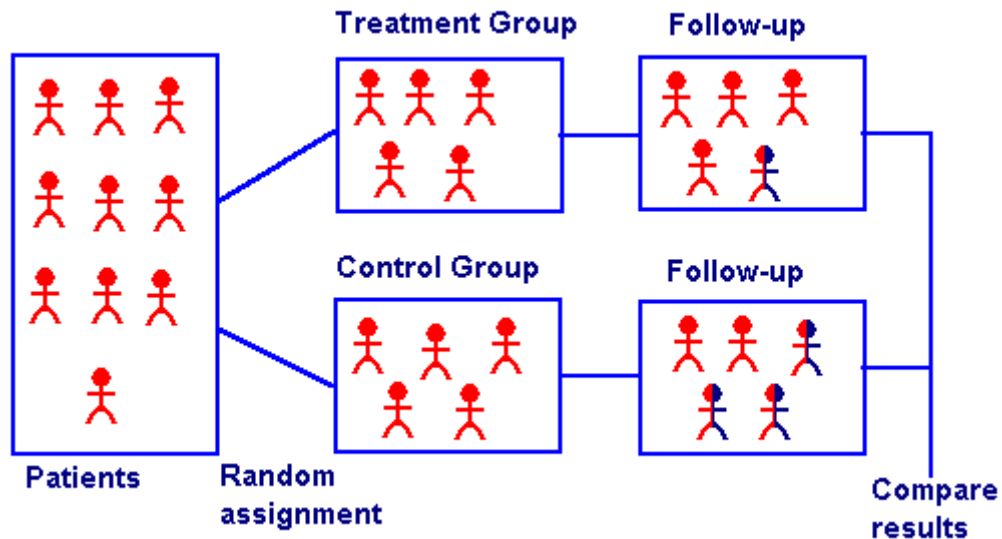
For example, studies of treatments that consist essentially of taking pills are very easy to do double blind - the patient takes one of two pills of identical size, shape, and color, and neither the patient nor the physician needs to know which is which.

A double blind study is the most rigorous clinical research design because, in addition to the randomization of subjects which reduces the risk of bias, it can eliminate the placebo effect which is a further challenge to the validity of a study.

The placebo effect could be thought of in this way:

1. Patients who believe they are receiving a new experimental treatment tend to be more optimistic about the outcome. This means that, when asked, they tend to minimize health problems and give more weight to positive effects. They also tend to take better care of themselves and comply better with the conditions of the experiment. There is also substantial evidence that, independent of all this, patients who have positive beliefs about their treatment do better than patients who do not. In many situations, the placebo effect is at least as strong as any objective effects of the treatment!
2. Doctors who believe that a patient is receiving a new experimental treatment tend to be more optimistic about that patient's chances, evaluate their state of health more favorably, and communicate positive expectations to the patients, who in turn try to get better so as to prove their doctor right!

Randomized Controlled Studies



A randomized controlled study is one in which:

1. There are two groups, one treatment group and one control group. The treatment group receives the treatment under investigation, and the control group receives either no treatment or some standard default treatment.
2. Patients are randomly assigned to all groups.

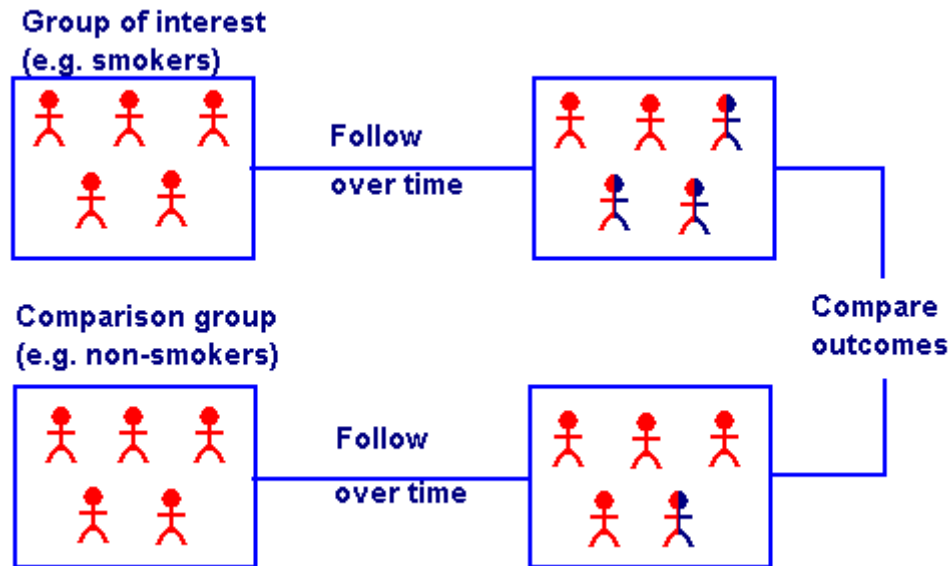
Assigning patients at random reduces the risk of bias and increases the probability that differences between the groups can be attributed to the treatment.

Having a control group allows us to compare the treatment with alternative choices. For instance, the statement that a particular medication cures 40% of cases tells us very little unless we also know how many cases get better on their own! (Or with a different treatment).

With certain research questions, randomized controlled studies cannot be done for ethical reasons. For instance, it would be unethical to attempt to measure the effect of smoking on health by asking one group to smoke two packs a day and another group to abstain, since the smoking group would be subject to unnecessary harm.

Randomized controlled trials are the standard method of answering questions about the effectiveness of different therapies. If you have a therapy question, first look for a randomized controlled trial, and only go on to look for other types of studies if you don't find one.

Cohort Studies



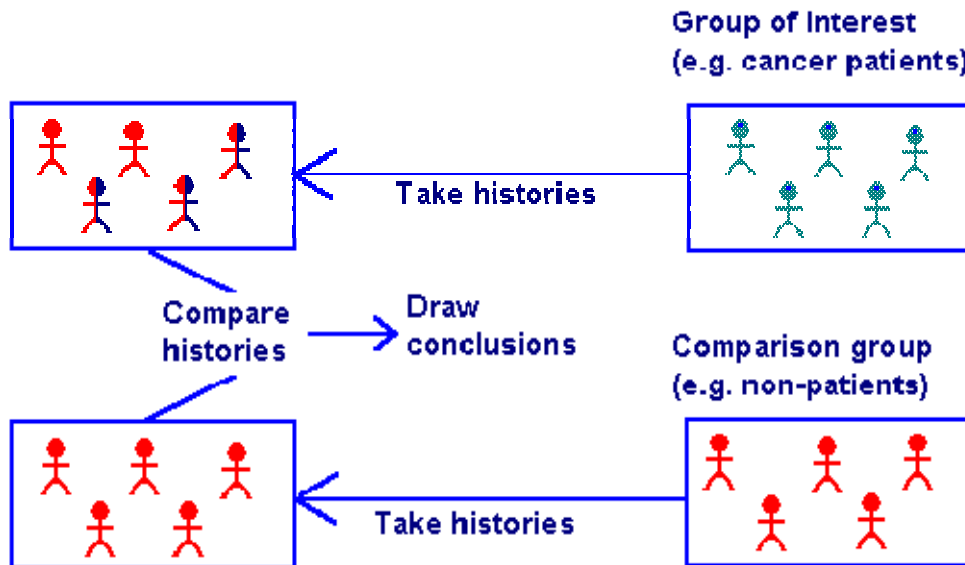
A Cohort Study is a study in which patients who presently have a certain condition and/or receive a particular treatment are followed over time and compared with another group who are not affected by the condition under investigation.

For instance, since a randomized controlled study to test the effect of smoking on health would be unethical, a reasonable alternative would be a study that identifies two groups, a group of people who smoke and a group of people who do not, and follows them forward through time to see what health problems they develop.

Cohort studies are not as reliable as randomized controlled studies, since the two groups may differ in ways other than in the variable under study. For example, if the subjects who smoke tend to have less money than the non-smokers, and thus have less access to health care, that would exaggerate the difference between the two groups.

The main problem with cohort studies, however, is that they can end up taking a very long time, since the researchers have to wait for the conditions of interest to develop. Physicians are, of course, anxious to have meaningful results as soon as possible, but another disadvantage with long studies is that things tend to change over the course of the study. People die, move away, or develop other conditions, new and promising treatments arise, and so on. Even so, cohort studies are generally preferred to case control studies, since they involve far fewer statistical problems and generally produce more reliable answers.

Case Control Studies



Case control studies are studies in which patients who already have a certain condition are compared with people who do not.

For example: a study on which lung cancer patients are asked how much they smoked in the past and the answers are compared with a sample of the general population would be a case control study.

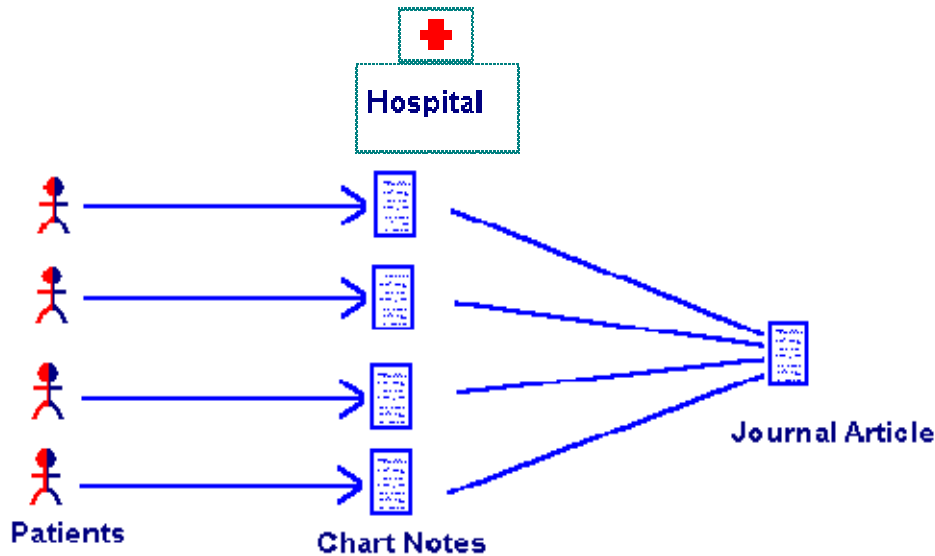
Case control studies are less reliable than either randomized controlled trials or cohort studies. Just because there is a statistical relationship between two conditions does not mean that one condition actually caused the other. For instance, lung cancer rates are higher for people without a college education (who tend to smoke more), but that does not mean that someone can reduce his or her cancer risk just by getting a college education.

The main advantages of case control studies are:

- They can be done quickly. By asking patients about their past history, researchers can quickly discover effects that otherwise would take many years to show themselves.
- Researchers don't need special methods, control groups, etc. They just take the people who show up at their institution with a particular condition and ask them a few questions.

The first study to suggest a new medical conclusion will often be a case control study, perhaps designed to check on a hypothesis suggested by a case series. If possible, researchers will generally try to confirm the results with a randomized controlled trial or a cohort study.

Case Series and Case Reports



Case series and case reports consist either of collections of reports on the treatment of individual patients, or of reports on a single patient.

For example: one of your patients has a condition that you have never seen or heard of before and you are uncertain what to do. A search for case series or case reports may reveal information that will assist in a diagnosis. However, for any reasonably well-known condition you should be able to get better evidence. Case series and case reports, since they use no control group with which to compare outcomes, have no statistical validity.