



YOUR PRACTICE / BUSINESS

Understanding Levels of Evidence

Martha Brown Menard, PhD, LMT

The concept of levels of evidence is a cornerstone of research literacy and a great starting point for understanding basic principles of how research works. Health care in the U.S. has embraced an evidence-based standard of care - health care treatments or interventions used need to be empirically tested, and demonstrated to be both safe and effective.

How Do We Know What We Know?

First, let's consider how we know what we know. There are at least five different ways of knowing: tradition, authority, trial and error, logical reasoning and finally, the scientific method.¹ The first two are based on taking someone else's word for it. The other three are increasingly more systematic ways of evaluating the accuracy of knowledge claims.

Tradition is based on the idea that certain practices continue just "because that's the way it's always been done." However, tradition inhibits the search for new approaches and may perpetuate ideas that are actually contradictory to the available evidence. Authority is based on the opinion of experts who have distinguished themselves in their fields. But all authorities are subject to error.

A good example of the problem with relying on tradition or authority is provided by [Dr. John Kellogg](#), an influential physician practicing massage in the early 1900s, who stated firmly that massage was contraindicated for people with cancer. For decades afterward, massage therapists accepted this statement as fact and avoided working with cancer patients.

Based on more current information, and on observational research,² we now know that massage is beneficial for reducing the physical discomfort and emotional stress associated with the treatment of cancer.



With trial and error, individuals try one approach and evaluate its effect. The process ends once a successful approach is found. The problem with trial and error is that potentially better solutions never get tested. Perhaps a patient has a sore trapezius muscle. They do nothing and it eventually gets better. Doing nothing works, but the patient will never know if something else might have worked better or more quickly.

In logical reasoning, individuals apply deductive or inductive reasoning to think their way through a situation. Deduction starts with a general premise, such as "heat increases circulation and relaxes muscle," and then applies the general premise to a specific situation, in this case, the [sore trapezius muscle](#). The opposite process occurs with inductive reasoning. Specific examples or cases are observed and a more general theory is then generated. The potential problem is the validity of a given conclusion based on observations can be limited by a small number of examples or a narrow range of cases observed.

The Scientific Method

This approach is the most rigorous way of acquiring new knowledge and is based on the idea of testing whether a hypothesis can be demonstrated to be true or false. We assume there is a single observable and measurable reality, one that we can all perceive and agree upon, and that there is a relationship between cause and effect. In order to understand how cause is related to effect, the scientist or investigator must ask questions and gather relevant information in a systematic way. The hypothesis or research question determines the kind of information necessary and the methods used to test it, whether is true or false.

The beauty of the scientific method is that it presents evidence which all observers can see. Another is its insistence that all knowledge is provisional. Because new information could be discovered at any time and show that what we thought was true is completely incorrect, scientific method works best by showing a hypothesis is false. We can only show a hypothesis has survived

repeated attempts to disprove it, and accept it with the caveat "as far as we know now." Once a hypothesis has been demonstrated to be false, however, it remains that way. The history of science is full of numerous examples - at one time everyone "knew" the Earth was flat.

Levels of Evidence

In many kinds of health care research, we want to know whether a certain treatment (cause) is responsible for a particular outcome (effect). For example, does chiropractic help premature babies gain weight? What factors (cause) contribute to developing an illness or promoting health (effect)? In critical evaluation of a research study reported in a journal article, we are assessing the evidence the authors/researchers have presented. How well have they argued their case that this treatment or factor is responsible for that result?

TRADITIONAL HIERARCHY OF RESEARCH STUDY DESIGNS

WEAKEST EVIDENCE

Case report

Case series

Correlational study

Before-and-after
treatment without control

Cross-sectional study

Case-control study

Before-and after
reatment with control

Cohort study

Randomized controlled
trial (RCT)

Meta-analysis

Systematic review

STRONGEST EVIDENCE

In any research article, the authors are presenting their findings and offering their interpretation of what those findings mean. In other words, they are making a case for the reader to decide to what extent the reader accepts, accepts with reservations or rejects the evidence presented, based on how strongly the authors have linked cause with effect.

Levels of evidence refers to the idea there are varying degrees or levels of evidence of cause and effect that different kinds of research studies can provide. It's rare for a single study to demonstrate a hypothesis so conclusively that it is beyond question. The more common scenario is that a body of evidence accumulates, study by study, with different studies contributing different degrees or kinds of evidence.

Within the research community, there is a hierarchy of weight based on the strength of the association between cause and effect that particular designs incorporate. The relative strength of each design is often shown schematically by an arrow running from top to bottom, with studies that provide weaker degrees of evidence at the top and the strongest designs at the bottom, based on the idea of a research pyramid.

All of the studies from case report down to the case-control study are considered descriptive or observational; that is, the investigator is describing or observing what is naturally occurring without interfering in any way, and then reporting what is seen. Descriptive and observational studies can be used to note associations between events and to generate hypotheses.

The before-and-after treatment with control, cohort, and randomized, controlled trials are examples of experimental designs. In these studies, the investigator is actively intervening by assigning some participants to receive a treatment and then comparing those participants to others who receive a different intervention (or no intervention), and reporting the results. Experimental studies provide a stronger link between cause and effect; in these designs, the investigator is systematically testing hypotheses.

Meta-analyses and systematic reviews are evaluations of groups of studies on a specific research question. By combining studies and then assessing them as a group, the authors can evaluate the weight of the evidence as a whole on that question.

The Circle of Methods

Another useful way to conceptualize the relative strengths and weaknesses of different research designs has been proposed by Walach, et al.³ Rather than a hierarchy, which tends to devalue designs that are ranked as lower in the cause-and-effect hierarchy, Walach and colleagues present a circle, in which experimental designs are balanced by more observational and descriptive ones.

These authors point out that the hierarchical model was founded and works well for pharmacological interventions, but has since been generalized to other, more complex interventions such as physical therapy, massage, acupuncture, and surgery, for which it is often inadequate and a poor fit for evaluating complex interventions.

Instead of holding up the RCT as a gold standard against which all other designs must be considered inferior, Walach's model proposes a multiplicity of methods that counterbalance the strengths and weaknesses of individual designs, and that can be considered equally rigorous, with the added benefit of greater clinical applicability. A similar model, called the evidence house, was previously proposed by Wayne Jonas.⁴

Experts on complementary and integrative therapies, many of whom are dually trained as researchers and practitioners, believe most complementary therapies are complex interventions. Factors such as the patient-practitioner relationship, and expectations, beliefs, attitudes and preferences seem likely to be moderating variables that can influence the specific effect of an intervention or mediate the biological mechanism through which an effect is produced. The circular model of levels of evidence offers a more helpful approach to evaluating the effectiveness of these therapies.

References

1. Portney LG & Watkins MP. *Foundations of Clinical Practice*. Inglewood Cliffs, NJ: Appleton & Lange, 1993; pp. 4-9.
2. Cassileth BR, Vickers AJ. Massage therapy for symptom control: outcome study at a major cancer center. *J Pain Symptom Manage*. 2004 Sep;28(3):244-9.
3. Jonas WB. Building an evidence house: challenges and solutions to research in complementary and alternative medicine. *Forsch Komplementarmed Klass Naturheilkd*, 2005 Jun;12(3):159-67.
4. Walach H, et al. Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC Med Research Methodol*, 2006, 6:29.

JULY 2016