

The ranking and reliability of evidence

We are constantly bombarded by information of all types—at what point does information become evidence?

Morley C. Sutter, MD, PhD

Important decisions often are based on evidence. Judges and juries assess evidence; so do physicians and scientists and so should politicians. For example, what is the evidence for the role of cholesterol in atherosclerosis? What is the evidence for global warming? But one person's evidence might be called a fairy tale by someone else. In each of these situations information is being examined. When is information evidence and when is it just a fairy tale? When does it lead to sound medicine and when to quackery?

The practice of medicine is both a science and an art. The art of medicine comes in to play when the science proves inadequate in a particular case. Presumably science—and therefore evidence-based medicine—is used wherever possible and is the first approach to diagnosis and treatment. In both the art and the science, information is used. If the art is used without previous consideration of the evidence, the medicine is liable to be inappropriate.

Evidence-based medicine has become a focal point in the practice of medicine. Several groups have published consensus rankings of the reliability of medical evidence derived from different types of studies. An early list of the hierarchy of medical studies

from least to most reliable was as follows: expert opinion, non-experimental studies, observational studies, non-randomized interventional studies, randomized controlled studies, and systematic reviews and meta-analysis of randomized controlled trials.¹ Subsequent rankings have incorporated the quality of individual studies into this list, but the ranking is the same, with expert opinion being least reliable.² It is the purpose of the present essay to broaden the look at reliability of evidence to go beyond medicine, to include everyday life.

What is information?

All evidence is based on information. Information exists in all of the forms of energy that can stimulate our five senses: sight, sound, touch, smell, and taste. Some would say that information also can come from within us as in hunches, gut feelings, intuition, dreams or visions, and as “extrasensory perception.” This essay, however, deals only with information in the form of words both written and spoken that originate outside of us. Such an approach seems justified inasmuch as words constitute the most important method of communication between people.

There are other forms of sensory-stimulating communication, such as body language, odors, and sign lan-

guage or other codes, but for effect, none has the impact, durability, or reproducibility of words, both spoken and written. Therefore conversations with colleagues and the medical literature are the most important sources of medical evidence.

When is information evidence?

All information constitutes evidence at some level of reliability. There is an old saying, “believe half of what you see and none of what you hear.” Clearly this referred to the days before photographic or electronic imaging. Now one almost could say, “believe nothing of what you see and none of what you hear,” so rampant and skilled is electronic sound and image modification. Computer-generated graphics have revolutionized animation and video games. Voice-over TV images are routine and lip-synch is an accepted art form. So what can be done to assess reliability of sound and sight in the form of words?

It is very difficult to verify the authenticity of any electronic source of words or images, so sophisticated is the technology used to alter electronic signals. This is illustrated by the publicity given to attempts to verify the

Dr Sutter is professor emeritus at the University of British Columbia.

voice of Osama bin Laden on a scratchy tape. There is a wonderful line in the movie *Wag the Dog* when Dustin Hoffman, playing the part of a Hollywood producer asks rhetorically, “Do you remember the image of the smart bomb striking its target in the Gulf War? Back lot” (of Hollywood). There also are people who still believe that humans never landed on the moon, that the images were faked.

It is almost impossible to assess information given by a skilled liar, but it can be equally difficult to judge the quality of information from someone who believes they are telling the truth when they are not. Often this is the case when religious matters are discussed. Then often rhetoric, enthusiasm, and belief systems prevail.

This essay is an attempt to provide a baseline or starting point for assessment of verbal information from a variety of sources. The scope is broader than the sources used in assessing reliability of information in evidence-based medicine. I hope what follows is of use in promoting rational discussion in the use of evidence in both medical and non-medical decision-making. A list of sources of verbal information is followed by comments on each source.

Information as evidence (from least- to most-reliable sources)

1. Advertisements (including political speeches)
2. Hearsay
3. Testimonials (including most Internet sites)
4. Testimonials from friends or authorities
5. Sworn testimony
6. Recorded observations
7. Recorded systematic observations
8. Recorded results of interventional experimentation with randomization and appropriate controls (including blinding of the observer(s) and subjects when necessary)
9. Recorded results from replications

of experiments as in 8

10. Successful predictions based on a model derived from recorded systematic observations (e.g., tide tables and eclipses)

11. Successful predictions based on a model derived from experimentation (e.g., atomic fission and fusion)

Only information derived from the sources referred to in points 8–11 allows causal relationships to be defined. Information from the sources indicated in points 1–7 can give clues but not conclusions.

1. Advertisements

Advertisements are designed to sell something, either oneself as in politics or a product as in commerce. Half-truths, innuendo, and incomplete information abound in advertising. Catch-phrases and attention-getting devices are prominent. The advertiser really is concerned only with getting you to pay attention, to sell, not to provide reliable information. Although there are organizations that promote concepts such as truth or ethics in advertising and there are laws against false advertising, the definitions of *truth* and *ethics* are elastic. Therefore advertisements never should be used as sources of information to be used as evidence. Advertisements concerning drugs are just that: advertisements that are subject to the foibles of advertising in general—no better, no worse. I leave it to the reader to decide what constitutes an advertisement for a drug.

2. Hearsay

Information garnered during casual conversation or overheard at a cocktail party represents hearsay as is gossip deliberately passed on. Urban myths such as finding alligators in the toilet bowl also are examples. Hearsay information is always suspect. Corridor consultations would fall in this category; they can be informative, but always should be carefully weighed.

3. Testimonials

Testimonials from individuals are a favorite means of promoting a product, either tangible or intangible. The former would include grown or manufactured products; the latter includes ideas and beliefs. Testimony springs from within the individual, who often makes a statement without the necessity of providing details. Whether the statement is true depends entirely on the awareness and integrity of the individual who makes it. This sort of testimonial can be classified as proclaiming a belief, either that something is true or that something occurred. The status of the person providing the testimonial often is used to lend credence to what is said. Testimonials are a form of advertising in which beliefs are being promulgated. Much of the information available on the Internet (World Wide Web) falls into this category.

4. Testimonials from friends, celebrities, or authority figures

Such testimonials have no more inherent reason to be reliable than ordinary testimonials but are widely used in advertising. Sports heroes and entertainers commonly are used to capture one’s attention. Word-of-mouth recommendations from friends also would fall into this category. Information from many lecturers, both academic and commercial (e.g., Deepak Chopra or Andrew Weil) is another example. The information in public presentations from such popular lecturers is often delivered skillfully, enthusiastically, and in an entertaining style but it is still just a testimonial. Expert opinion is named as the least reliable source of evidence in the ranking of reliability of sources for use in evidence-based medicine.^{1,2}

5. Sworn testimony

Sworn testimony is an attempt by the legal system to allow statements of a witness or declarations of a concerned party to be used as evidence. Sworn

Continued on page 18

Continued from page 17

testimony is a key part of the rule of law and is necessary for a civilized society to function. Therefore sworn affidavits and notarized statements commonly are used to ensure that statements are factual both in business and in bureaucratic or quasi-legal circumstances such as pensions and legacies. The penalties for lying while

Sworn testimony is an improvement on a simple testimonial but still is susceptible to error that might or might not be detected.

under oath are severe and the court provides means whereby persons may be cross-examined to attempt to verify or falsify the statements of witnesses or other participants in legal proceedings. But people do lie under oath or their observations can be mistaken. Sworn testimony is an improvement on a simple testimonial but still is susceptible to error that might or might not be detected.

6. Recorded observations

Recorded observations are a common source of information, particularly historical. A good example would be a personal diary or log of daily activities.

Case reports, both spoken and published, are medical examples. Such information provides insight as to what is happening, but its accuracy is imperfect; what is recorded is entirely dependent on the observer and what he or she is interested in at that moment. Blogs (web logs) are a good example and are notorious for inconsistency, eccentricity, and egocentricity.

7. Recorded systematic observations

More reliable are systematic observations, as when astronomers record the position of the stars at the same time each day. Other examples are Darwin's records of the species of finches in the Galapagos or a mariner recording the time of high and low tide each day. The important features are that the record be written down and that the observations be made in a consistent and systematic way over a prolonged period. Once having been made, the observations often are used to build models, either mathematical or physical (i.e., mechanical, electronic, or chemical). The models are then tested as in seeing whether a type of airship flies or whether the models can accurately predict tides or eclipses.

The heavens in particular have been observed over thousands of years and systematic records kept for almost as long. Even if they are accurate, the context of the records, their interpretation, and how they are used is important. For instance, Copernicus interpreted his observations to mean that the earth went around the sun and not the sun around the earth. In 1610, when Galileo agreed with his fellow astronomer Copernicus, Galileo was tried as a heretic. The concept that the earth was not the centre of the universe did not sit well with the leaders of the Catholic church of that time. The context was inappropriate for them and the Church apologized for their error only in 1990.

Much of the research in medicine falls into this category of systematic

observation including anatomical, physiological, biochemical, and pharmacodynamic studies. Epidemiological studies are also of this type and often identify correlations between phenomena. Observations without interventions can only disprove a prediction or hypothesis. That is, if no correlation is found between two variables it is most unlikely that one causes the other. Observational studies are the sort most likely to be reported by the news media: "A particular behavior or phenomenon is linked to the incidence of disease X." An example is linking abortion to breast cancer. The word "cause" is never used but the implication is that the abortion somehow caused breast cancer. Such observational studies of associations must be viewed with great skepticism; they can only give clues to causal relationships. Interventional studies where a variable is deliberately altered or there is successful predictive testing of a model (see below) are necessary to demonstrate causal relationships between phenomena.^{3,4}

8. Recorded results of interventional experimentation that is appropriately designed and carried out

Results of interventional experimentation allow choices to be made between ways of doing things. An example is how best to treat disease where randomized, controlled trials compare one treatment with a placebo or another treatment. The experiments must be designed appropriately and performed diligently for conclusions to be valid. Important points in design are randomization of treatment groups and their treatment; unbiased observers who do not know which treatment is being given to a particular patient plus patients who do not know what treatment they are receiving (double-blinding); accurate and appropriate endpoints of the success or failure of treatment and appropriate comparisons (e.g., with placebo or best available

treatment). Above all, the appropriate controls must be used so that like is compared with like and every attempt must be made to avoid bias in selection of subjects. The statistics applied to such clinical experimentation were originally developed by R.A. Fisher for testing various interventions to improve crop yields at an agricultural research station in the UK. Results from randomized, controlled trials are the gold standard of evidence-based medicine.

9. Repetition and confirmation of the experiments as in 8

Replication of results is an important aspect of science that is sometimes forgotten. Although there can be statistical variability, similar experiments should yield similar results. If the results differ between experiments, reasons for the difference should be sought. Systematic reviews (meta-analyses) of randomized, controlled trials are an attempt to quantify the degree of reproducibility of interventional experiments.

10. Successful predictions based on a model derived from recorded systematic observations

Being able to predict eclipses and the movements of the tides are remarkable achievements that took centuries to develop. Detailed observations were made and recorded and many models, both mechanical (“music of the spheres”) and mathematical were made and tested empirically until accurate predictions could be made. The current prediction of tides includes something like 60 variables and the digital computer has been a boon. The description of gravity by Newton and the realization that the moon exerted gravity was an important step in understanding tides, but so were fluid dynamics as described by la Place. The point is that interventional experimentation could not be done, but extremely accurate and successful predictive models

resulted from the systematic observations and testing of various models of both the heavens and the seas.

Nothing in medicine can be predicted with the accuracy of tides or eclipses, for our understanding of bodily function is very superficial; our models are inadequate for accurate prediction. Even when the model developed by a biologist leads to Nobel Prizes as in the description of how a nerve impulse is conducted (ion fluxes), the double helix of DNA, or how blood vessels dilate by producing nitric oxide, the described model is brilliant, but still inadequate for all but simple predictions of biological behavior.

11. Successful predictions or outcomes based on a model derived from experimentation

Theory together with results from earlier experimentation had led to Fermi’s experimental production of a nuclear chain reaction. This strongly suggested that the then current models of the atom were substantially correct. The ultimate, dramatic test of theory was to produce an atomic explosion. When the atomic bomb exploded at Los Alamos, further proof of the structure of the atom and of the mathematical model, $E=mc^2$ was obtained. As stated previously, this degree of understanding of a model is not present in any system of biology; such systems are very complex and our knowledge quite superficial. Therefore physician-scientists cannot predict the future behavior of a biological system except in an approximate way; biological systems are more complex than those that produce even atomic explosions. The success in achieving an atomic explosion, however, also changed how science is funded and organized. It led to the development of the National Institutes of Health in the US and to the system of applying for research grants that we have today. But that is another story.

Epilogue

Human behavior is complex; much of

it is learned by rote and almost automatic or intuitive, without conscious thought. The hierarchy of evidence as discussed above would not apply to decisions made on that basis but only to decisions made on an intellectual basis. It might be useful, however if more of our decisions were made in the latter context, particularly the important ones.

References

1. United States Department of Health and Human Services. Agency for Health Care Policy and Research. Acute pain management: Operative or medical procedures and trauma. Rockville, MD: AHCPR, 1993:107. Clinical practice guideline No. 1, AHCPR publication No. 92-0023.
2. Harbour R, Miller J. Scottish Intercollegiate Guidelines Network Grading Review Group, A new system for grading recommendations in evidence based guidelines. *BMJ* 2001;323:334-336.
3. Sutter MC. Assigning causation in disease: Beyond Koch’s postulates. *Perspect Biol Med* 1996;39:581-592.
4. Sutter MC. To know: The need for science. *BC Med J* 2000;42:338-340.