

Education and debate

Effect of interpretive bias on research evidence

Ted J Kaptchuk

Doctors are being encouraged to improve their critical appraisal skills to make better use of medical research. But when using these skills, it is important to remember that interpretation of data is inevitably subjective and can itself result in bias.

Facts do not accumulate on the blank slates of researchers' minds and data simply do not speak for themselves.¹ Good science inevitably embodies a tension between the empiricism of concrete data and the rationalism of deeply held convictions. Unbiased interpretation of data is as important as performing rigorous experiments. This evaluative process is never totally objective or completely independent of scientists' convictions or theoretical apparatus. This article elaborates on an insight of Vandenbroucke, who noted that "facts and theories remain inextricably linked ... At the cutting edge of scientific progress, where new ideas develop, we will never escape subjectivity."² Interpretation can produce sound judgments or systematic error. Only hindsight will enable us to tell which has occurred. Nevertheless, awareness of the systematic errors that can occur in evaluative processes may facilitate the self regulating forces of science and help produce reliable knowledge sooner rather than later.

Interpretative processes and biases in medical science

Science demands a critical attitude, but it is difficult to know whether you have allowed for too much or too little scepticism. Also, where is the demarcation between the background necessary for making judgments (such as theoretical commitments and previous knowledge) and the scientific goal of being objective and free of preconceptions? The interaction between data and judgment is often ignored because there is no objective measure for the subjective components of interpretation. Taxonomies of bias usually emphasise technical problems that can be fixed.³ The biases discussed below, however, may be present in the most rigorous science and are obvious only in retrospect.

Quality assessment and confirmation bias

The quality of any experimental findings must be appraised. Was the experiment well performed and are the outcomes reliable enough for acceptance? This scrutiny, however, may cause a confirmation bias: researchers may evaluate evidence that supports their prior belief differently from that apparently challeng-

ing these convictions. Despite the best intentions, everyday experience and social science research indicates that higher standards may be expected of evidence contradicting initial expectations.

Two examples might be helpful. Koehler asked 297 advanced university science graduate students to evaluate two supposedly genuine experiments after being induced with different "doses" of positive and negative beliefs through false background papers.⁴ Questionnaires showed that their beliefs were successfully manipulated. The students gave significantly higher rating to reports that agreed with their manipulated beliefs, and the effect was greater among those induced to hold stronger beliefs. In another experiment, 398 researchers who had previously reviewed experiments for a respected journal were unknowingly randomly assigned to assess fictitious reports of treatment for obesity. The reports were identical except for the description of the intervention being tested. One intervention was an unproved but credible treatment (hydroxycitrate); the other was an implausible treatment (homoeopathic sulphur). Quality assessments were significantly higher for the more plausible version.⁵ Such confirmation bias may be common.^{w1 w2}



Harvard Medical School, Osher Institute, 401 Park Drive, Boston, MA 02215, USA

Ted J Kaptchuk
assistant professor of medicine

ted_kaptchuk@hms.harvard.edu

BMJ 2003;326:1453-5



Further references (denoted by the prefix "w") are available on bmj.com

Definitions of interpretation biases

Confirmation bias—evaluating evidence that supports one's preconceptions differently from evidence that challenges these convictions

Rescue bias—discounting data by finding selective faults in the experiment

Auxiliary hypothesis bias—introducing ad hoc modifications to imply that an unanticipated finding would have been otherwise had the experimental conditions been different

Mechanism bias—being less sceptical when underlying science furnishes credibility for the data

"Time will tell" bias—the phenomenon that different scientists need different amounts of confirmatory evidence

Orientation bias—the possibility that the hypothesis itself introduces prejudices and errors and becomes a determinate of experimental outcomes

Expectation and rescue and auxiliary hypothesis biases

Experimental findings are inevitably judged by expectations, and it is reasonable to be suspicious of evidence that is inconsistent with apparently well confirmed principles. Thus an unexpected result is initially apt to be considered an indication that the experiment was poorly designed or executed.^{6 w3} This process of interpretation, so necessary in science, can give rise to rescue bias, which discounts data by selectively finding faults in the experiment. Although confirmation bias is usually unintended, rescue bias is a deliberate attempt to evade evidence that contradicts expectation.

Instances of rescue bias are almost as numerous as letters to the editors in journals. The avalanche of letters in response to the Veterans Administration Cooperative randomised controlled trial examining the efficacy of coronary artery bypass grafting published in 1977 is a well documented example.⁷ The trial found no significant difference in mortality between 310 patients treated medically and 286 treated surgically. A subgroup of 113 patients with obstruction of the left main coronary artery, however, clearly benefited from surgery.⁸ Instead of settling the clinical question, the trial spurred fierce debate in which supporters and detractors of the surgery perceived flaws that, they claimed, would skew the evidence away from their preconceived position. Each stakeholder found selective faults to justify pre-existing positions that reflected their disciplinary affiliations (cardiology *v* cardiac surgeon), traditions of research (clinical *v* physiological), and personal experience.⁹

Auxiliary hypothesis bias is a form of rescue bias. Instead of discarding contradictory evidence by seeing fault in the experiment, the auxiliary hypothesis introduces ad hoc modifications to imply that an unexpected finding would have been otherwise had the experimental conditions been different. Because experimental conditions can easily be altered in so many ways, adjusting a hypothesis is a versatile tool for saving a cherished theory.^{w4} Evidence pointing to an unwelcome finding in a randomised controlled trial, for example, can easily be dismissed by arguments against the therapeutic dose, its timing, or how patients

were selected. Lakatos termed such reluctance to accept an experimental verdict a scientist's "thick skin."¹⁰ Thus, when early randomised controlled trials showed that hormone replacement therapy did not reduce the risk of coronary heart disease,¹¹ advocates of hormone replacement therapy argued that it was still valuable for primary prevention because the study group was women with established coronary heart disease, making the disease too far advanced to benefit from the treatment.

Plausibility and mechanism bias

Evidence is more easily accepted when supported by accepted scientific mechanisms. This understandable tendency to be less sceptical when underlying science furnishes credibility can give rise to mechanism bias. Often, such scientific plausibility underlies and overlaps the other biases I've described. Many examples exist where with hindsight it is clear that plausibility caused systematic misinterpretation of evidence. For example, the early negative evidence for hormone replacement therapy would have undoubtedly been judged less cautiously if a biological rationale had not already created a strong expectation that oestrogens would benefit the cardiovascular system.^{12 w5} Similarly, the rationale for antiarrhythmic drugs for myocardial infarction was so imbedded that each of three antiarrhythmic drugs had to be proved harmful individually before each trial could be terminated.^{13 w6} And the link between *Helicobacter pylori* and peptic ulcer was rejected initially because the stomach was considered to be too acidic to support bacterial growth.¹⁴

Waiting for more evidence and "time will tell" bias

The position that more evidence is necessary before making a judgment indicates a judicious attitude that is central to a scientific scepticism. None the less, different scientists seem to need different amounts of confirmatory evidence to feel satisfied. This discrepancy in duration conceals a subjective process that easily can become a "time will tell" bias. The evangelist, at one extreme, is quick to accept the data as good evidence (or even proof). Evangelists often have a vested intellectual, professional, or personal commitment and may have taken part in the experiment being assessed. At the other extreme are the snails, who invariably find the data unconvincing, perhaps because of their personal and intellectual investment in old "facts." At the two extremes, as well as at all points in between, there is no objective way to tell whether good judgment or systematic error is operating. Max Planck described the "time will tell" bias cynically: "a new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it."¹⁵

Hypothesis and orientation bias

The above categories of potential biases all occur after data are collected. Sometimes, however, conviction may affect the collection of data, creating orientation bias. Psychologists call this the "experimenter's hypothesis as

Summary points

Evidence does not speak for itself and must be interpreted for quality and likelihood of error

Interpretation is never completely independent of a scientist's beliefs, preconceptions, or theoretical commitments

On the cutting edge of science, scientific interpretation can lead to sound judgment or interpretative biases; the distinction can often be made only in retrospect

Common interpretative biases include confirmation bias, rescue bias, auxiliary hypothesis bias, mechanism bias, "time will tell" bias, and orientation bias

The interpretative process is a necessary aspect of science and represents an ignored subjective and human component of rigorous medical inquiry

an unintended determinant of experimental results.¹⁶ Thus, psychology graduate students, when informed that rats were specially bred for maze brightness, found that these rats outperformed those bred for maze dullness, despite both groups really being standard laboratory rats assigned at random.¹⁷ Somehow, experimental and recording errors tend to be larger and more in the direction supporting the hypothesis.^{17, 18}

Numerous studies have noted that randomised controlled trials sponsored by the pharmaceutical industry consistently favour new therapies.¹⁸ Research outcomes seem to be affected by what the researcher is looking for. It is unclear to what extent these apparent successes are the result of publication bias or matters of study design. Nonetheless, such results are consistent with an orientation bias and explain the fact that some early double blind randomised controlled trials performed by enthusiasts show efficacy—like hyperbaric oxygen for multiple sclerosis¹⁹ or endotoxin antibodies for Gram negative septic shock²⁰—whereas subsequent trials cannot replicate the outcome.¹⁹

Comments

This article is written from the perspective of philosophy of science. From a statistical point of view, the arguments presented are obviously compatible with a subjectivist or bayesian framework that formally incorporates previous beliefs in calculations of probability. But even if we accept that probabilities measure objective frequencies of events, the arguments still apply. After all, the overall experiment still has to be assessed.

I have argued that research data must necessarily undergo a tacit quality control system of scientific scepticism and judgment that is prone to bias. Nonetheless, I do not mean to reduce science to a naive relativism or argue that all claims to knowledge are to be judged equally valid because of potential subjectivity in science. Recognition of an interpretative process does not contradict the fact that the pressure of additional unambiguous evidence acts as a self regulat-

ing mechanism that eventually corrects systematic error. Ultimately, brute data are coercive. However, a view that science is totally objective is mythical, and ignores the human element of medical inquiry. Awareness of subjectivity will make assessment of evidence more honest, rational, and reasonable.²¹

This article is a shortened version of a paper written for a seminar on bias led by Fredrick Mosteller at Harvard University and reflects his helpful feedback. Peter Goldman criticised earlier versions of the article and helped make it understandable. The comments of Iain Chalmers and Al Fishman have been helpful, as was the dedicated research of Cleo Youtz. All errors and shortcomings of the paper belong solely to the author.

Funding: In part from grants IR01 AT00402-01 and IR01 AT001414 from the National Institutes of Health, Bethesda, MD. Competing interests: None declared.

- 1 Elstein AS. Human factors in clinical judgment: discussion of Scriven's clinical judgment. In : Engelhardt HT, Spicker SF, Towers B. *Clinical judgment: a critical appraisal*. Dordrecht: Reidel, 1979.
- 2 Vandembroucke JP. Medical journals and the shaping of medical knowledge. *Lancet* 1998;352:2001-6.
- 3 Sackett DL. Bias in analytic research. *J Chron Dis* 1979;32:51-63.
- 4 Koehler JJ. The influence of prior beliefs on scientific judgments of evidence quality. *Organ Behav Hum Decision Processes* 1993;56:28-55.
- 5 Resch KI, Ernst E, Garrow J. A randomized controlled study of reviewer bias against an unconventional therapy. *J R Soc Med* 2000;93:164-7.
- 6 Kuhn TS. *The structure of scientific revolutions*. Chicago: University of Chicago Press, 1962.
- 7 Jones DS. Visions of cure: visualization, clinical trials, and controversies in cardiac therapeutics, 1968-1998. *Isis* 2000;91:504-41.
- 8 Murphy ML, Hultgren HN, Detre K, Thomsen J, Takaro T, Veterans Administration Cooperative Study. Treatment of chronic stable angina: a preliminary report of survival data of the randomized Veterans Administration cooperative study. *N Engl J Med* 1977;297:621-7.
- 9 Special correspondence: a debate on coronary bypass. *N Engl J Med* 1977;297:1464-70.
- 10 Lakatos I. *The methodology of scientific research programmes*. Cambridge: Cambridge University Press, 1978.
- 11 Herrington DM, Rebuossin DM, Brosnihan B, Sharp PC, Shumaker SA, Snyder TE, et al. Effects of estrogen replacement on the progression of coronary-artery atherosclerosis. *N Engl J Med* 2000;343:522-9.
- 12 Nabel EG. Coronary heart disease in women—an ounce of prevention. *N Engl J Med* 2000;343:572-574.
- 13 Cardiac Arrhythmia Suppression Trial II Investigators. Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. *N Engl J Med* 1992;327:227-33.
- 14 Thagard P. *How scientists explain disease*. Princeton: Princeton University Press, 1999.
- 15 Planck M. *Scientific autobiography and other papers*. London: Williams and Norgate, 1950.
- 16 Rosenthal R. On the social psychology of the psychological experiment: the experimenter's hypothesis as an unintended determinant of experimental results. *Am Sci* 1963;51:268-83.
- 17 Rosenthal R, Fode KL. Three experiments in experimenter bias. *Psychol Rep Monog* 1963;3:12-8.
- 18 Djulbegovic B, Lacey M, Cantor A, Fields KK, Bennett CL, Adams J, et al. The uncertainty principle and industry-sponsored research. *Lancet* 2000;356:635-8.
- 19 Robinson I. Clinical trials and the collective ethic: the case of hyperbaric oxygen therapy and the treatment of multiple sclerosis. In: Weisz G, ed. *Social science perspectives on medical ethics*. Philadelphia: University of Pennsylvania Press, 1991.
- 20 Ziegler EJ, Fisher CJ Jr, Sprung CL, Straube RC, Sadoff JC, Foulke GE, et al. Treatment of gram-negative bacteremia and septic shock with HA-1A human monoclonal antibody against endotoxin. A randomized, double-blind, placebo controlled trial. *N Engl J Med* 1991;324:429-36.
- 21 Toulmin S. *Return to reason*. Cambridge: Harvard University Press, 2001. (Accepted 19 March 2003)

Endpiece

In praise of neurotics

All the greatest things we know have come to us from neurotics. It is they and they only who have founded religions and created great works of art. Never will the world be conscious of how much it owes to them, nor above all what they have suffered in order to bestow their gifts on it.

Marcel Proust, *Guermites Way*, 1921