

[Home](#)[Contents](#)jameslindlibrary.org

Avoiding biased comparisons:

Differences in the way treatment outcomes are assessed

Using blinding to reduce bias in assessing treatment outcomes

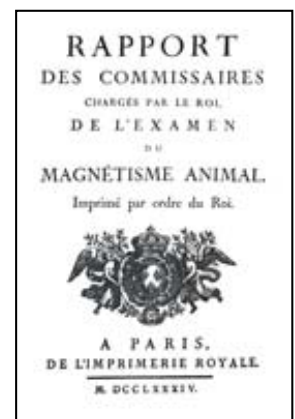
For some outcomes used to assess treatment – survival, for example - biased assessment is very unlikely because there is little room for opinion. This was the case in some of the 18th century tests of surgical procedures, where survival was the main measure of treatment success or failure ([Faure 1759](#)). The assessment of most other outcomes, however, either always involves subjectivity (as with patients' symptoms), or may involve it. The biases that lead to these misperceptions are termed observer biases. They cause a particular problem when people believe that they already 'know' the effect of a treatment, or when they may have particular reasons for preferring one of the treatments being compared. When measures are not taken to reduce biased outcome assessments in treatment comparisons, treatment effects tend to be overestimated (Schulz et al. 1995). The greater the element of subjectivity in assessing outcomes, the greater the need to reduce these observer biases to ensure fair tests of treatments.

In these common circumstances, 'blinding' of patients and doctors is a desirable element of fair tests. The earliest blinded (masked) assessment of a treatment appears to have been performed by a commission of inquiry appointed by Louis XVI in 1784 to investigate Anton Mesmer's claims of the effects of 'animal magnetism' ([Commission Royale 1784](#)).

The commission assessed whether the purported effects of this new healing method were due to any 'real' force, or due to the 'illusions of the mind'. Blindfolded people were told that they were receiving or not receiving magnetism when in fact, at times, the reverse was happening. The people being studied felt the effects of 'animal magnetism' only when they were told they were receiving the treatment, but not otherwise (Kaptchuk 1998; Schulz et al. 2002).

Using placebos to achieve blinding

A few years after the tests of the effects of animal magnetism, John Haygarth conducted an experiment using a sham device (a placebo) to achieve blinding ([Haygarth 1800](#)). The cartoon that accompanies this paragraph shows a doctor treating a wealthy client with a device patented and marketed by Elisha Perkins. Perkins claimed that his '[tractors](#)' – small metal rods - cured a variety of ailments through 'electrophysical force'. In a pamphlet entitled '*Of the imagination as a cause and as a cure of disorders of the body: exemplified by fictitious tractors*', John Haygarth reported how he put Perkins' claims to a fair test. In a series of patients who were unaware of the details of his evaluation, he used a cross-over study to compare the patented, metal tractors (which were meant to work through 'electrophysical force') with wooden 'tractors' that looked identical ('placebo tractors'). He was unable to detect any benefit of the metal tractors ([Haygarth 1800](#)).



John Haygarth's fair test of Perkins' tractors is an early example of the use of placebos to achieve blinding to reduce biases in assessing the outcome of treatments. Placebos became a research tool in the debates on homeopathy, the nineteenth century's other major form of unconventional healing. Homeopaths often used blind assessment and placebo controls for their "provings", which tested the effects of their remedies on healthy volunteers ([Löhner 1835](#); Kaptchuk 1998). One of the most sophisticated placebo-controlled tests took place under the Milwaukee Academy of Medicine in 1879-1880. This trial was 'double-blind': both patients and experimenters were kept unaware as to whether the treatment was a genuine homeopathic remedy or a sugar pill ([Storke et al. 1880](#)).

It was not until much later that a more skeptical attitude in mainstream medicine led to a recognition that there was a need to adopt blinded assessment and placebos to assess the validity of its own claims. Inspired principally by pharmacologists, German researchers gradually adopted masked assessment. For example, in 1918, Adolf Bingel reported that he had tried to be "as objective as possible" when comparing two different treatments for diphtheria ([Bingel 1918](#)). He assessed whether he or his colleagues could guess which patients had received which treatment: "I have not relied on my own judgment alone, but have sought the views of the assistant physicians of the diphtheria ward, without informing them about the nature of the serum under test.



Their judgment was thus completely without prejudice. I am keen to see my observations checked independently, and most warmly recommend this 'blind' method for the purpose." ([Bingel 1918](#)). In fact, no difference was detected between the two treatments. A strong tradition of blind assessment developed in Germany, and this was codified by the clinical pharmacologist Paul Martini ([Martini 1932](#)).

Blind assessment in the modern English-speaking world first began when pharmacologists were influenced by the German tradition, as well as by an indigenous 'quackbuster' movement that used masked assessment (Kaptchuk 1998). By the 1930's, they had taken the lead in using placebo controls in clinical experiments. For example, two of the UK Medical Research Council's earliest fair tests were of treatments for the common cold. It would have been very difficult to interpret their results had 'double blinding' not been used to prevent patients and doctors knowing which patients had received the new drugs and which had received placebos ([MRC 1944](#); [MRC 1950](#)). Harry Gold's strenuous advocacy of the importance of blinded assessment appears to have had a particularly important influence in the United States ([Conference on Therapy 1954](#)).

Blinding observers when it is impossible to blind patients and clinicians

Sometimes it is simply impossible to blind patients and doctors to the identity of the treatments being compared, for example, when surgical treatments are compared with drug treatments, or no treatment. Even in these circumstances, however, steps can be taken to reduce biased assessment of treatment outcomes. Independent observers can be kept unaware of which treatments have been received by which patients. For example, in the early 1940s a test compared patients with pulmonary tuberculosis receiving the then standard treatment - bed rest - with other patients who received, in addition, injections of the drug streptomycin. The researchers felt that it would be unethical to inject inactive placebos in patients allocated to bed rest alone simply to achieve 'blinding' of the patients and doctors treating them ([MRC 1948](#)), but they took alternative precautions to reduce biased assessment of outcomes. Although there was little danger of biased assessment of the principal outcome (survival), subjectivity could have biased the assessment of the chest X-rays. Accordingly, X-rays were assessed by doctors who were kept unaware of whether they were evaluating outcome in a patient who had been treated with streptomycin or one treated with bed rest alone.

Together with randomization, masked assessment, when possible using placebos, has now become one of the crucial methodological components of fair tests of treatments.

Cite as: Editorial commentary (2007). Differences in the way treatment outcomes are assessed. The James Lind Library (www.jameslindlibrary.org).

Show JLL records: illustrating [control of observer bias](#)

Next essay: [Interpreting unbiased comparisons](#)

Select other essay:

References

Bingel A (1918). Über Behandlung der Diphtherie mit gewöhnlichem Pferdeserum. Deutsches Archiv für Klinische Medizin 125:284-332.

Commission Royale (1784). Rapport des commissaires chargés par le roi du magnetisme animal. Paris: Imprimerie royale.

Conference on Therapy (1954). How to evaluate a new drug. American Journal of Medicine 17:722-727.

Faure (1759). Recueil des pieces qui ont concouru pour le prix de L'Académie Royale de Chirurgie. Vol 8. Paris, P. Al Le Prieur.

Haygarth J (1800). Of the imagination, as a cause and as a cure of disorders of the body: exemplified by fictitious tractors, and epidemical convulsions. Bath: R. Crutwell.

Kaptchuk TJ (1998). Intentional ignorance: a history of blind assessment and placebo controls in medicine. Bulletin of the History of Medicine 72:389-433.

Löhner G (1835), on behalf of a Society of truth-loving men. Die Homoöopathischen Kochsalzversuche zu

Nürnberg [The homeopathic salt trials in Nuremberg].

Martini P (1932). Methodenlehre der Therapeutischen Untersuchung. Berlin: Springer.

Medical Research Council (1944). Clinical trial of patulin in the common cold. *Lancet* 2: 373-375.

Medical Research Council (1948). Streptomycin treatment of pulmonary tuberculosis: a Medical Research Council investigation. *BMJ* 2: 769-782.

Medical Research Council (1950). Clinical trials of antihistaminic drugs in the prevention and treatment of the common cold. *BMJ* 2: 425-431.

Schulz KF, Chalmers I, Hayes RJ, Altman DG (1995). Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 273: 408-412.

Schulz KF, Chalmers I, Altman D (2002). The landscape and lexicon of blinding. *Annals of Internal Medicine* 136: 254-259.

Storke EF, Martin R, Rosenkrans EM, Ford J, Schloemilch A, McDermott GC, Carlson OW (1880). Final report of the Milwaukee test of the thirtieth dilution. *Homeopathic Times: A Monthly Journal of Medicine, Surgery and the Collateral Sciences* 7: 280-281.

[Home](#)

[Contents](#)

[Comments welcome](#)