James Lind Library

Illustrating the development of fair tests of treatments in health care

1.0 Introduction to JLL Explanatory Essays

Cite as: The James Lind Library 1.0 Introduction to JLL Explanatory Essays (http://jameslindlibrary.org/essays/1-0-introduction-to-jll-explanatory-essays/)

> Despite acting with the best of intentions, health professionals have sometimes done more harm than good to the patients who have looked to them for help. Some of this suffering can be reduced by ensuring that fair tests are done to address uncertainties about the effects of treatments.

Over the past half century, health care has had a substantial impact on people's chances of living longer and being free of serious health problems. It has been estimated that health care has been responsible for between a third and a half of the increase in life expectancy and an average of five additional years free of chronic health problems. Even so, the public could have obtained – and still could obtain – far better value for the very substantial resources its invests in research intended to improve health (www.researchwaste.net). Furthermore, some of the treatment disasters of the past could have been prevented, and others could be prevented in future.

Misleading claims about the effects of treatments are common, so all of us should understand how valid claims about the effects of treatments are made. Without this knowledge, we risk concluding that useless treatments are helpful, or that helpful treatments are useless. The James Lind Library has been created to improve general understanding of fair tests of treatments in health care, and how these have evolved over time.

The Explanatory Essays in The James Lind Library have been written to promote wider understanding of why fair tests of treatments are needed, and what they have come to consist of. You can access each essay by clicking on the relevant links below; or, if you want to download all of the essays so that they can be printed out together for reading off screen, click here.

Fair tests of treatments

Biases

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

1.1 Why treatment uncertainties should be addressed

Cite as: The James Lind Library 1.1 Why treatment uncertainties should be addressed (http://jameslindlibrary.org/essays/1-1-why-treatment-uncertainties-should-be-addressed/)

Ignoring uncertainties about the effects of treatments has led to avoidable suffering and deaths. To reduce this suffering and premature mortality, treatment uncertainties must be acknowledged and addressed, first by reviewing systematically what is already known, and then in well designed research to reduce continuing uncertainties.

Trying to do more good than harm

Why do we need fair tests of treatments in health care? Have not doctors, for centuries, 'done their best' for their patients? Sadly, there are many examples of doctors and other health professionals harming their patients because treatment decisions were not informed by what we consider now to be reliable evidence about the effects of treatments. With hindsight, health professionals in most if not all spheres of health care have harmed their patients inadvertently, sometimes on a very wide scale (Silverman 2003; Grimes 2007). Indeed, patients themselves have sometimes harmed other patients when, on the basis of untested theories and limited personal experiences, they have encouraged the use of treatments that have turned out to be harmful. The question is not whether we must blame these people, but whether the harmful effects of inadequately tested treatments can be reduced. They can, to a great extent.

Acknowledging that treatments can sometimes do more harm than good is a prerequisite for reducing unintended harm (Gregory 1772; Haygarth 1800; Fordyce 1802; Behring 1893). We then need to be more ready to admit uncertainties about treatment effects, and to promote tests of treatments to adequately reduce uncertainties. Such tests are fair tests.

Why theories about the effects of treatments must be tested in practice

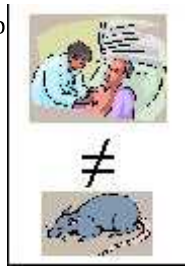
People have often been harmed because treatments have been based only on theories about how health problems should be treated, without testing how the theories played out in practice. For example, for centuries people believed the theory that illnesses were caused by 'humoral imbalances', and patients were bled and purged, made to vomit and take snuff, in the belief that this would end the supposed imbalances. Still, as long ago as the 17th century, a lone Flemish doctor was impertinent enough to challenge the medical authorities of the time to assess the validity of their theories by proposing a fair test of the results of their unpleasant treatments (Van Helmont 1648).

By the beginning of the 19th century, British military surgeons had begun to show the harmful effects of bloodletting for treating "fevers" (Robertson 1804; Lesassier Hamilton 1816). A few decades later, the practice was also challenged by a Parisian physician (Louis 1835). Yet at the beginning of the 20th century, orthodox practitioners in Boston, USA, who were not using bloodletting to treat pneumonia were still being judged negligent (Silverman 1980). Indeed, Sir William Osler, one of the most influential medical authorities in the world, who was generally cautious about recommending unproven treatments, advised his readers at the end of the 19th century that: "during the last decades we have certainly bled too little. Pneumonia is one of the diseases in which a timely venesection [bleeding] may save life. To be of service it should be done

early. In a full-blooded, healthy man with a high fever and bounding pulse the abstraction of from twenty to thirty ounces of blood is in every way beneficial" (Osler 1892).

Although the need to test the validity of theories in practice was recognized by some people at least a millennium ago (Ibn Hindu 10th-11th century), this important principle is still too often ignored. For instance, based on untested theory, Benjamin Spock, the influential American child health expert, informed the readers of his best selling book 'Baby and Child Care' that a disadvantage of babies sleeping on their backs was that, if they vomited, they would be more likely to choke. Dr Spock therefore advised his millions of readers to encourage babies to sleep on their tummies (Spock 1966). We now know that this advice, apparently rational in theory, led to the cot (crib) deaths of tens of thousands of infants (Gilbert et al. 2005).

The use of drugs to prevent heart rhythm abnormalities in people having heart attacks provides another example of the dangers of applying untested theory in practice. Because heart rhythm abnormalities are associated with an increased risk of early death after heart attack, the theory was that drugs that reduced rhythm abnormalities would also reduce early deaths. Just because a theory seems reasonable doesn't mean that it is necessarily right, however. Years after the drugs had been licensed and adopted in practice, it was discovered that they actually increase the risk of sudden death after heart attack. Indeed, it has been estimated that, at the peak of their use in the late 1980s, they may have been killing as many as 70,000 people every year in the United States alone (Moore 1995) – many more than the total number of Americans who died in the Vietnam War.



On the other hand, misplaced confidence in theoretical thinking as a guide to practice has also resulted in some treatments being rejected inappropriately because researchers did not believe that they could work. Theories based on the results of animal research, for example, sometimes correctly predict the results of treatment tests in humans, but this is not always the case (Perel et al. 2007). Based on the results of experiments in rats, some researchers became convinced that there was no point in giving clot-dissolving drugs to patients who had experienced heart attacks more than six hours previously. Had no such patients participated in some of the fair tests of these drugs we would not know that they can benefit from this treatment (Fibrinolytic Therapy Trialists' Collaborative Group 1994).

Observations in clinical practice or in laboratory and animal research may suggest that particular treatments will or will not benefit patients; but as these and many other examples make clear, it is essential to use fair tests to find out whether, in practice, these treatments do more good than harm, or vice versa.

< Previous Essay | Next Essay >

References

Behring, Boer, Kossel H (1893). Zur Behandlung diphtheriekranker Menschen mit Diphtherieheilserum. Deutsche Medicinische Wochenschrift 17:389-393.

Fibrinolytic Therapy Trialists' Collaborative Group (1994). Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomised trials of more than 1000 patients. Lancet 1994;343:311-322.

Fordyce G (1802). Â A second dissertation on fever. London: J Johnson

Gilbert R, Salanti G, Harden M, See S (2005). Infant sleeping position and the sudden infant death syndrome:

systematic review of observational studies and historical review of recommendations from 1940 to 2002. International Journal of Epidemiology 34:874-87.

Gregory J (1772). Lectures on the duties and qualifications of a physician. London: Strahan and Cadell.

Grimes DA (2007). Discovering the need for randomized controlled trials in obstetrics: a personal odyssey. JLL Bulletin: Commentaries on the history of treatment evaluation (www.jameslindlibrary.org).

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.

Ibn Hindu (10th-11th century CE; 4th-5th century AH). Miftah al-tibb wa-minhaj al-tullab [The key to the science of medicine and the students' guide].

Lesassier Hamilton A (1816). Dissertatio Medica Inauguralis De Synocho Castrensi (Inaugural medical dissertation on camp fever). Edinburgh: J Ballantyne.

Louis PCA (1835). Recherches sur les effets de la saignèe dans quelques maladies inflammatoires et sur l'action de l'èmètique et des vèsicatoires dans la pneumonie. Paris: Librairie de l'Acadèmie royale de mèdecine.

McPherson K (2004). Where are we now with hormone replacement therapy? BMJ 328:357-358.

Moore TJ (1995). Deadly Medicine. New York: Simon and Schuster.

Osler W (1892). Principles and Practice of Medicine. London: Appleton, p 530.

Perel P, Roberts I, Sena E, Wheble P, Briscoe C, Sandercock S, Macleod M, Mignini LE, Jayaram P, Khan KS (2007). Comparison of treatment effects between animal experiments and clinical trials: systematic review. BMJ 334:197-200.

Robertson R (1804). Observations on the diseases incident to seamen, 2nd edn. Vol. 1, London: for the author.

Silverman W (1980). In: Chalmers I, McIlwaine G (eds). Perinatal Audit and Surveillance. London: Royal College of Obstetricians and Gynaecologists, 1980:110.

Silverman WA (2003). Personal reflections on lessons learned from randomized trials involving newborn infants, 1951 to 1967. JLL Bulletin: Commentaries on the history of treatment evaluation (www.jameslindlibrary.org).

Spock B (1966). Baby and Child Care. 165th printing. New York: Pocket Books, pp 163-164.

Van Helmont JB (1648). Ortus medicinæ: Id est Initia physicæ inaudita. Progressus medicinae novus, in morborum ultionem, ad vitam longam [The dawn of medicine: That is, the beginning of a new Physic. A new advance in medicine, a victory over disease, to (promote) a long life]. Amsterodami: Apud Ludovicum Elzevirium, pp 526-527.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

1.2 Why treatment comparisons are essential

Cite as: The James Lind Library 1.2 Why treatment comparisons are essential (http://jameslindlibrary.org/essays/1-2-why-treatment-comparisons-are-essential/)

Treatment comparisons are required to take account of the natural course of health problems, placebo effects, and to go beyond impressions about treatment effects. But treatment comparisons need to be fair to avoid untrustworthy and sometimes dangerously incorrect conclusions about the effects of treatments.

Is a treatment better than nature and time?

Patients and healthcare professionals hope that treatments will be helpful. These optimistic expectations can have a very positive effect on everybody's satisfaction with health care, as the British doctor Richard Asher noted in one of his essays for doctors:

> "If you can believe fervently in your treatment, even though controlled tests show that it is quite useless, then your results are much better, your patients are much better, and your income is much better too. I believe this accounts for the remarkable success of some of the less gifted, but more credulous members of our profession, and also for the violent dislike of statistics and controlled tests which fashionable and successful doctors are accustomed to display." (Asher 1972)

People often recover from illness without any specific treatment: nature and time are great healers. As Oliver Wendell Holmes suggested in the 19th century when there were very few useful treatments (Holmes 1861), "I firmly believe that if the whole materia medica, as now used, could be sunk to the bottom of the sea, it would be all the better for mankind – and all the worse for the fishes." The progress and outcome of illness if left untreated must obviously be taken into account when treatments are being tested: treatment may improve or it may worsen outcomes. Writers over the centuries have drawn attention to the need to be sceptical about claims that the effects of treatments can improve on the effects of nature (list of records coded Principles of Testing). Put another way, "If you leave a dose of 'flu to nature, you'll probably get over it in a week; but if you go to the doctor, you'll recover in a mere seven days."

Placebo effects

In the knowledge that much illness is self-limiting, doctors sometimes prescribe inert treatments in the hope that their patients will derive psychological benefit – the so-called placebo effect. Patients who believe that a treatment will help to relieve their symptoms – even though the treatment, in fact, has no physical effects – may well feel better. Doctors have recognized the importance of using placebos for centuries (list relevant records). For example, William Cullen referred to his use of a placebo as long ago as 1772 (Cullen 1772), and references to placebos increased during the 19th century (Cummings 1805; Ministry of Internal Affairs 1832; Forbes 1846). Because Austin Flint believed that orthodox drug treatment was usurping the credit due to 'nature', he gave thirteen patients with rheumatism a 'placeboic remedy' consisting of a highly dilute extract of the bark of the quassia tree. The result was that "the favourable progress of the cases was such as to secure for the remedy generally the entire confidence of the patients" (Flint 1863). At Guy's Hospital in

London, William Withey Gull came to similar conclusions after treating 21 rheumatic fever patients "for the most part with mint water" (Sutton 1865). At the beginning of the 20th century William Rivers discussed psychologically-mediated effects of treatments in detail (Rivers 1908).

The need for comparisons

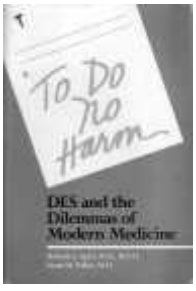
Just as the healing power of nature and the placebo effect have been recognized for centuries, so also has the need for comparisons to assess the effects of treatments over and above natural and psychologically-mediated effects. Sometimes treatment comparisons are made in people's minds: they have an impression that they or others are responding differently to a new treatment compared with previous responses to treatments. For example, Ambroise Paré, a French military surgeon, concluded that treatment of battle wounds with boiling oil (as was common practice) was likely to be harmful. He concluded this when the supply of oil ran out and his patients recovered more quickly than usual (Paré 1575). Most of the time, impressions like this need to be followed up by formal investigations, perhaps initially by analysis of healthcare records. Such impressions may then lead to carefully conducted comparisons. The danger arises when impressions alone are used as a guide to treatment recommendations and decisions.

Dramatic effects and moderate effects of treatments

Treatment comparisons based on impressions, or relatively restricted analyses, only provide reliable information in therare circumstances when treatment effects are dramatic (click here to list relevant records; Glasziou et al. 2007). The James Lind Library contains illustrations both of dramatic beneficial effects of treatments – for example, opium for pain relief (Tibi 2005), insulin for diabetes (Banting et al. 1922), liver diet for pernicious anaemia (Minot and Murphy 1926), sulpha drugs for infection after childbirth (Colebrook and Purdie 1937) and streptomycin for tuberculous meningitis (MRC 1948) – and of dramatic harmful effects, for example limb reduction deformities caused by thalidomide (McBride 1961). Sometimes a treatment, sulphonamide drugs, for example, can have a dramatic effect in some diseases, but modest or little effect in others (Loudon 2002). Most medical treatments don't have dramatic effects, however, and unless care is taken to avoid biased comparisons, dangerously mistaken conclusions about the effects of treatment may result.

Comparing treatments given today with treatments given in the past

It was partly because of reliance on biased comparisons with past experience that doctors and women believed that the drug diethylstilboestrol (DES) would reduce the risk of miscarriages and stillbirths. There was never any evidence from fair (unbiased) tests that DES could do this, and it was later shown that it caused cancer in the daughters of some of the pregnant women for whom it had been prescribed. A treatment that has not been reliably shown to be useful should not be promoted.



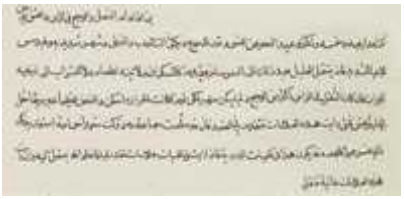
Comparing treatments given today with treatments given in the past only rarely provides a secure basis for a fair test (Behring et al. 1893; Roux et al. 1894), because relevant factors other than the treatments themselves change over time. For example, miscarriages and stillbirths are more common in first pregnancies than in later pregnancies. Comparing the frequency of miscarriages and stillbirths in later pregnancies in which DES was prescribed with the outcome of first pregnancies in which the drug wasn't used is thus likely to be a seriously misleading basis for assessing its effects. If possible, therefore, comparisons should involve giving different treatments at more or less the same time.

Comparing treatments in crossover tests in individual patients

Sometimes giving different treatments at more or less the same time may involve giving a patient different treatments one after the other – a so-called crossover test (Martini 1932;

click here to list relevant records). Sometime this is done in a single patient – a so-called N-of-1 trial. An early example of a crossover test was reported in 1786 by Dr Caleb Parry in Bath, England. He wanted to find out whether there was any reason to pay for expensive, imported Turkish rhubarb as a purgative for treating his patients, rather than using rhubarb grown locally in England. So he 'crossed-over' the type of rhubarb given to each individual patient at different times and then compared the symptoms each patient experienced while eating each type of rhubarb (Parry 1786). (He didn't find any advantage of the expensive rhubarb!) Treatment comparisons within individual patients have their place when their condition returns after stopping treatment. There are many circumstances in which this doesn't apply. For example, it is usually impossible to compare different surgical operations in this way, or treatments given for progressive conditions.

Comparing groups of patients given different treatments concurrently



Treatments are usually tested by comparing groups of people who receive different treatments. A comparison of two treatments will be unfair if relatively well people have received one treatment and relatively ill people have received the other, so the experiences of similar groups of people who receive different treatments over the same period of time must be compared. Al-Razi recognized this more than a thousand years ago when, wishing to reach a conclusion about how to treat patients with signs of early meningitis, he treated one group of patients and intentionally withheld treatment from a comparison group (al-Razi 10th century).

Sometimes studies are done to compare two or more treatments given separately or together (factorial trials) Comparisons with nature or with other treatments are needed for fair tests of treatments. If these comparisons are to be fair, they must address genuine uncertainties, avoid biases and the play of chance, and be interpreted carefully.

< Previous Essay | Next Essay >

References

al-Razi (10th century CE; 4th Century AH). Kitab al-Hawi fi al-tibb [The comprehensive book of medicine].

Asher R (1972). Talking sense. London: Pitman Medical.

Banting FG, Best CH, Collip JB, Campbell WR, Fletcher AA (1922). Pancreatic extracts in the treatment of diabetes mellitus. Canadian Medical Association Journal 12:141-146.

Behring, Boer, Kossel H (1893). Zur Behandlung diphtheriekranker Menschen mit Diphtherieheilserum. Deutsche Medicinische Wochenschrift 17:389-393.

Colebrook L, Purdie AW (1937). Treatment of 106 cases of puerperal fever by sulphanilamide. Lancet 2:1237-1242 & 1291-1294.

Cullen W (1772). Clinical lectures. Edinburgh, February-April, 218-9.

Cummings R (1805). Medical and Physical Journal, page 6.

Flint A (1863). A contribution toward the natural history of articular rheumatism; consisting of a report of thirteen cases treated solely with palliative measures. American Journal of the Medical Sciences 46:17-36.

Forbes J (1846). Homeopathy, allopathy and 'young physic.' British and Foreign Medical Review 21:225-265.

Glasziou P, Chalmers I, Rawlins N, McCulloch P (2007). When are randomised trials unnecessary? Picking

signal from noise.BMJ 334:349-351.

Holmes OW (1861). Currents and countercurrents in medical science. In: Works, 1861 Vol ix, p 185.

Loudon I (2002). The use of historical controls and concurrent controls to assess the effects of sulphonamides, 1936-1945. JLL Bulletin: Commentaries on the history of treatment evaluation (www.jameslindlibrary.org).

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

McBride WG (1961). Thalidomide and congenital abnormalities. Lancet 2:1358.

Medical Research Council (1948). Streptomycin treatment of tuberculous meningitis. Lancet 1:582-596.

Ministry of Internal Affairs (1823). [Conclusion of the Medical Council regarding homeopathic treatment]. Zhurnal Ministerstva Vnutrennih del, 3:49-63.

Minot GR, Murphy WP (1926). Treatment of pernicious anaemia by a special diet. JAMA 87:470-476.

Paré A (1575). Les oeuvres de M. Ambroise Paré conseiller, et premier chirugien du Roy avec les figures & portraicts tant de l'Anatomie que des instruments de Chirugie, & de plusieurs Monstres. Paris: Gabriel Buon.

Parry CH (1786). Experiments relative to the medical effects of Turkey Rhubarb, and of the English Rhubarbs, No. I and No. II made on patients of the Pauper Charity. Letters and Papers of the Bath Society III: 431-453.

Rivers WHR (1908). The influence of alcohol and other drugs on fatigue. London:Edward Arnold.

Roux E, Martin L, Chaillou A (1894). Trois cent cas de diphthérie traité par le serum antidiphthérique. Annales de l'Institut Pasteur 8:640-661.

Sutton HG (1865). Cases of rheumatic fever, treated for the most part by mint water. Collected from the clinical books of Dr Gull, with some remarks on the natural history of that disease. Guy's Hospital Report 11:392-428.

Tibi S (2005). The medicinal use of opium in ninth-century Baghdad. Leiden: Brill.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

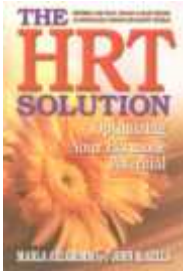
1.3 Why treatment comparisons must be fair

Cite as: The James Lind Library 1.3 Why treatment comparisons must be fair (http://jameslindlibrary.org/essays/1-3-why-treatment-comparisons-must-be-fair/)

> Untrustworthy treatment comparisons are those in which biases, or the play of chance, or both result in misleading estimates of the effects of treatments. Fair treatment comparisons avoid biases and reduce the effects of the play of chance.

Failure to test theories about treatments in practice is not the only preventable cause of treatment tragedies. Tragedies have also occurred because the tests used to assess the effects of treatments have been unreliable and misleading. The principles of fair tests have been evolving for at least a millennium (list records coded Principles of Testing) – and they continue to evolve today (Savovic et al. 2012; Jefferson et al. 2014).

For example, in the 1950s, theory and poorly controlled tests yielding unreliable evidence suggested that giving a synthetic sex hormone, diethylstilboestrol (DES), to pregnant women who had previously had miscarriages and stillbirths would increase the likelihood of a successful outcome of later pregnancies. Although fair tests had suggested that DES was useless, theory and the unreliable evidence, together with aggressive marketing, led to DES being prescribed to millions of pregnant women over the next few decades. The consequences were disastrous: some of the daughters of women who had been prescribed DES developed cancers of the vagina, and other children had other health problems, including malformations of their reproductive organs and infertility (Apfel and Fisher 1984).



Problems resulting from inadequate tests of treatments continue to occur. Again, as a result of unreliable evidence and aggressive marketing, millions of women were persuaded to use hormone replacement therapy (HRT), not only because it could reduce unpleasant menopausal symptoms, but also because it was claimed that it would reduce their chances of having heart attacks and strokes. When these claims were assessed in fair tests, the results showed that, far from reducing the risks of heart attacks and strokes, HRT increases the risks of these life-threatening conditions, as well as having other undesirable effects (McPherson 2004).

These examples of the need for fair tests of treatments are a few of many that illustrate how treatments can do more harm than good. Improved general knowledge about fair tests of treatments is needed so that – laced with a healthy dose of scepticism – we can all assess claims about the effects of treatments more critically. That way, we will all become more able to judge which treatments are likely to do more good than harm.

Fair tests entail taking steps to reduce the likelihood that we will be misled by the effects of biases of various sorts. Those addressed in the James Lind Library include design bias, allocation bias, co-intervention bias, observer bias, analysis bias, biases in assessing unanticipated effects, reporting bias, biases in systematic reviews, and researcher biases and fraud.

Essays on taking account of the play of chance address recording and interpreting numbers, quantifying uncertainty, and reducing the play of chance using meta-analysis.

< Previous Essay | Next Essay >

References

Apfel RJ, Fisher SM (1984). To do no harm: DES and the dilemmas of modern medicine. New Haven, Ct: Yale University Press.

Jefferson T, Jones MA, Doshi P, Del Mar CB, Hama R, Thompson MJ, Spencer EA, Onakpoya I, Mahtani KR, Nunan D, Howick J, Heneghan CJ (2014). Neuraminidase inhibitors for preventing and treating influenza in healthy adults and children. Cochrane Database of Systematic Reviews 2014, Issue 4. Art. No.: CD008965. DOI:10.1002/14651858.CD008965.pub4.

McPherson K (2004). Where are we now with hormone replacement therapy? BMJ 328:357-358.

Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, Als-Nielsen B, Balk EM, Gluud C, Gluud LL, Ioannidis JPA, Schulz KF, Beynon R, Welton NJ, Wood L, Moher D, Deeks JJ, Sterne JAC (2012). Influence of reported study design characteristics on intervention effect estimates from randomized controlled trials. Annals of Internal Medicine 157:429-438.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

2.0 Avoiding biased treatment comparisons

Cite as: The James Lind Library 2.0 Avoiding biased treatment comparisons (http://jameslindlibrary.org/essays/2-0-avoiding-biased-treatment-comparisons/)

What are biases? Biases in tests of treatments are those influences and factors that can lead to conclusions about treatment effects that are systematically different from the truth.

Sometimes treatments have dramatic effects (click here to list relevant records). These may be unintended and specific, for example, when a person has an allergic reaction to an antibiotic drug. Treatments can also have striking beneficial effects, like adrenaline for life-threatening allergic reactions (McLean-Tooke et al. 2003). Such striking effects are rare, however. Usually, treatment effects are more modest, but nevertheless well worth knowing about, for example, using aspirin to reduce risk of heart attack (Elwood 2004).

Aspirin doesn't prevent all premature deaths after a heart attack, but it does reduce the likelihood of death by about twenty per cent, which is important in such a common condition. If these moderate but important effects of most treatments are to be detected reliably, care must be taken to ensure that biased comparisons don't lead us to believe that treatments are useful when they are useless or harmful, or useless when they can actually be helpful.

Biases in tests of treatment are those influences and factors that can lead to conclusions about treatment effects that are systematically different from the truth. Although many kinds of biases can distort the results of health research (Sackett 1979), we have considered design bias, allocation bias, co-intervention bias, observer bias, analysis bias, biases in assessing unanticipated effects, reporting bias, biases in systematic reviews, and research biases and fraud.

Usually, the unfair tests of treatment resulting from these biases are not recognised for what they are. However, people with vested interests sometimes exploit these biases so that treatments are presented as if they are better than they really are (Sackett and Oxman 2003).

Whether biases are inadvertent or deliberate, the consequences are the same: unless tests of treatment are fair, some useless or harmful treatments will seem to be useful, while some useful treatments will seem useless or harmful.

< Previous Essay | Next Essay >

References

Elwood P (2004). The first randomised trial of aspirin for heart attack and the advent of systematic overviews of trials. The James Lind Library (www.jameslindlibrary.org).

McLean-Tooke APC, Bethune CA, Fay AC, Spickett GP (2003). Adrenaline in the treatment of anaphylaxis: what is the evidence? BMJ 327:1332-1335.

Sackett DL (1979). Bias in analytic research. Journal of Chronic Diseases 32:51-63.

Sackett DL, Oxman AD (2003). HARLOT plc: an amalgamation of the world's two oldest professions. BMJ 2003;327:1442-1445.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

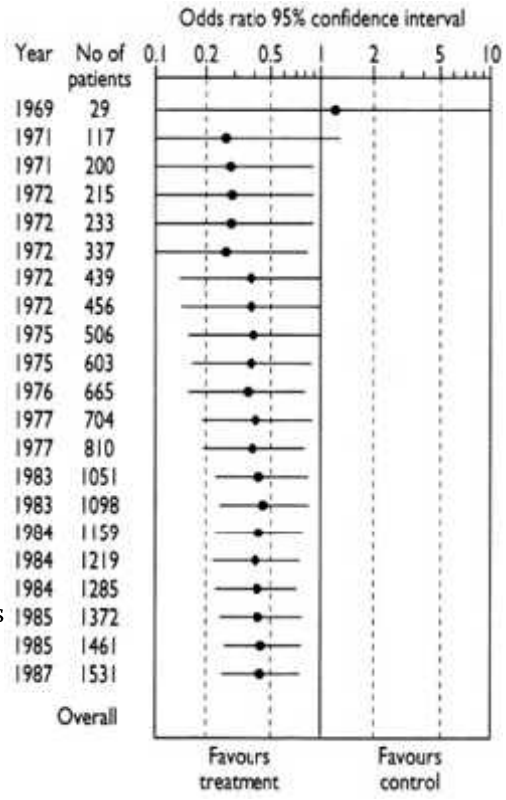
2.1 Why comparisons must address genuine uncertainties

Cite as: The James Lind Library 2.1 Why comparisons must address genuine uncertainties (http://jameslindlibrary.org/essays/2-1-why-comparisons-must-address-genuine-uncertainties/)

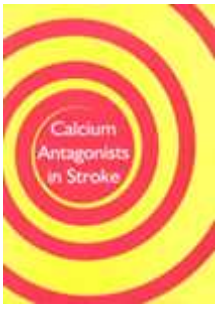
> The design of treatment research often reflects commercial and academic interests; ignores relevant existing evidence; uses comparison treatments known in advance to be inferior; and ignores needs of users of research results (patients, health professionals and others).

A good deal of research is done even when there are no genuine uncertainties. Researchers who fail to conduct systematic reviews of past tests of treatments before embarking on further studies sometimes don't recognise (or choose to ignore the fact) that uncertainties about treatment effects have already been convincingly addressed. This means that people participating in research are sometimes denied treatment that could help them, or given treatment likely to harm them.

The diagram that accompanies this and the following paragraph shows the accumulation of evidence from fair tests done to assess whether antibiotics (compared with inactive placebos) reduce the risk of post-operative death in people having bowel surgery (Lau et al. 1995). The first fair test was reported in 1969. The results of this small study left uncertainty about whether antibiotics were useful – the horizontal line representing the results spans the vertical line that separates favourable from unfavourable effects of antibiotics. Quite properly, this uncertainty was addressed in further tests in the early 1970s.



As the evidence accumulated, however, it became clear by the mid-1970s that antibiotics reduce the risk of death after surgery (the horizontal line falls clearly on the side of the vertical line favouring treatment). Yet researchers continued to do studies through to the late 1980s. Half the patients who received placebos in these later studies were thus denied a form of care which had been shown to reduce their risk of dying after their operations. How could this have happened? It was probably because researchers continued to embark on research without reviewing existing evidence systematically. This behaviour remains all too common in the research community, partly because some of the incentives in the world of research – commercial and academic – do not put the interests of patients first (Chalmers 2000).



Patients and participants in research can also suffer because researchers have not systematically reviewed relevant evidence from animal research before beginning to test treatments in humans. A Dutch team reviewed the experience of over 7000 patients who had participated in tests of a new calcium-blocking drug given to people experiencing a stroke. They found no evidence to support its increasing use in practice (Horn and Limburg 2001). This made them wonder about the quality and findings of the animal research that had led to the research on patients. Their review of the animal studies revealed that these

had never suggested that the drug would be useful in humans (Horn et al. 2001).

The most common reason that research does not address genuine uncertainties is that researchers simply have not been sufficiently disciplined to review relevant existing evidence systematically before embarking on new studies. Sometimes there are more sinister reasons, however. Researchers may be aware of existing evidence, but they want to design studies to ensure that their own research will yield favourable results for particular treatments. Usually, but not always, this is for commercial reasons (Djulbegovic et al. 2000; Sackett and Oxman 2003; Chalmers and Glasziou 2009; Macleod et al. 2014). These studies are deliberately designed to be unfair tests of treatments. This can be done by withholding a comparison treatment known to help patients (as in the example given above), or giving comparison treatments in inappropriately low doses (so that they don't work so well), or in inappropriately high doses (so that they have more unwanted side effects) (Mann and Djulbegovic 2012). It can also result from following up patients for too short a time (and missing delayed effects of treatments), and by using outcome measures ('surrogates') that have little or no correlation with the outcomes that matter to patients.

It may be surprising to readers of this essay that the research ethics committees established during recent decades to ensure that research is ethical have done so little to influence this research malpractice. Most such committees have let down the people they should have been protecting because they have not required researchers and sponsors seeking approval for new tests to have reviewed existing evidence systematically (Savulescu et al. 1996; Chalmers 2002). The failure of research ethics committees to protect patients and the public efficiently in this way emphasizes the importance of improving general knowledge about the characteristics of fair tests of medical treatments.

< Previous Essay | Next Essay >

References

Chalmers I. Current Controlled Trials: an opportunity to help improve the quality of clinical research. Current Controlled Trials in Cardiovascular Medicine 2000;1:3-8. Available: http://cvm.controlled-trials.com/content/1/1/3

Chalmers I (2002). Lessons for research ethics committees. Lancet 359:174.

Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet 2009;374:86-89. doi:10.1016/S0140-6736(09)60329-9.

Djulbegovic B, Lacevic M, Cantor A, Fields KK, Bennett CL, Adams JR, Kuderer NM, Lyman GH (2000). The uncertainty principle and industry-sponsored research. Lancet 356:635-638.

Horn J, Limburg M (2001). Calcium antagonists for acute ischemic stroke (Cochrane Review). In: The Cochrane Library, Issue 3, Oxford: Update Software.

Horn J, de Haan RJ, Vermeulen M, Luiten PGM, Limburg M (2001). Nimodipine in animal model experiments of focal cerebral ischaemia: a systematic review. Stroke 32:2433-38.

Lau J, Schmid CH, Chalmers TC (1995). Cumulative meta-analysis of clinical trials builds evidence for exemplary clinical practice. Journal of Clinical Epidemiology 48:45-57.

Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JPA, Salman RA-S, Chan A-W, Glasziou P. Biomedical research: increasing value, reducing waste. Lancet 2014;383:4-6.

Mann H, Djulbegovic B (2012). Comparator bias: why comparisons must address genuine uncertainties. JLL Bulletin: Commentaries on the history of treatment evaluation (www.jameslindlibrary.org)

Sackett DL, Oxman AD (2003). HARLOT plc: an amalgamation of the world's two oldest professions. BMJ 2003;327:1442-1445.

Savulescu J, Chalmers I, Blunt J (1996). Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. BMJ 313:1390-1393.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

2.2 The need to compare like-with-like in treatment comparisons

Cite as: The James Lind Library 2.2 The need to compare like-with-like in treatment comparisons (http://jameslindlibrary.org/essays/2-2-the-need-to-compare-like-with-like-in-treatment-comparisons/)

Allocation bias results when treatment comparisons fail to ensure that, apart from the treatments being compared, 'like will be compared with like'.

Comparing different treatments given to groups of people

Treatment comparisons usually entail comparing the experiences of groups of people who have received different treatments. If these comparisons are to be fair, the composition of the groups must be similar – so that like will be compared with like. If those who receive one treatment are more likely anyway to do well (or badly) than those receiving an alternative treatment, this allocation bias makes it impossible to be confident that outcomes reflect differential effects of the treatments, rather than the effects of nature and the passage of time.

The 18th century surgeon William Cheselden was aware of the 'dissimilar groups' problem when surgeons were comparing their respective mortality rates after operations to remove bladder stones. Cheselden pointed out that it was important to take account of the ages of the people treated by different surgeons. He drew attention to the fact that mortality rates varied with the patients' ages (Cheselden 1740) – older patients were more likely than younger patients to die. This meant that, if one wished to compare the frequency of deaths in groups of patients who had undergone different types of operation, one had to take account of differences in the ages of the patients in the comparison groups.

Comparing the experiences and outcomes of patients who happened to have received different treatments in the past is still used today as a way of trying to assess the effects of treatments. The challenge is to know whether the comparison groups were sufficiently alike before receiving treatment. This is illustrated by attempts to assess the effects of hormone replacement therapy (HRT) by comparing the illness experiences of women who had used HRT with those of other women who had not used it. As subsequent analysis of fair tests of HRT showed, trying to assess the effects of treatments in retrospect in this way can sometimes be dangerously misleading (McPherson 2004).

It is rarely possible to be completely confident that comparison groups selected from people who have been given one treatment in the past are comparable in all the respects that matter with people who have more recently received an alternative treatment. This is the case even if some information about the patients who have received different treatments is available (such as their ages, or their past history of illness). Other information that may be of great importance (such as the likelihood of spontaneous recovery) may simply not be available.

A better approach is to plan the treatment comparisons before starting treatment. For example, before beginning his comparison of six treatments for scurvy on board HMS Salisbury in 1747, James Lind took care to select patients who were at a similar stage of this often fatal disease. He also ensured that they had the same basic diet and were accommodated in similar conditions. These were factors, other than treatment, that might have influenced their likelihood of recovering (Lind 1753). Comparable efforts must be made to try to ensure that treatment comparison groups are composed of similar people.

Unbiased assembly of treatment comparison groups using alternation or randomisation

Although Lind took care to ensure that the sailors in his six comparison groups were alike, he didn't describe how he decided which sailors would receive which of the six treatments. There is only one way to ensure that treatment comparison groups are set up in such a way that they are similar in all the ways that matter, known and unknown. This is by using some form of chance process to assemble treatment comparison groups, so avoiding biased selection for different treatments before starting treatment.

One hundred years after Lind, an army doctor, Graham Balfour, illustrated how this could be done in a test to see whether belladonna prevented scarlet fever in children. In the military orphanage for which he was responsible, he used alternation – "to prevent the imputation of selection" – to decide which boys would receive and which would not receive belladonna (Balfour 1854). Alternation is one of several unbiased methods for assembling similar treatment comparison groups before giving the treatments being compared. During the first half of the 20th century, there are many examples of treatment comparison groups being assembled using alternation or rotation (for example MRC 1944), or by drawing lots (Colebrook 1929) – for example, using dice (Doull et al. 1931), coloured beads (Theobald 1937), or random sampling numbers (Bell 1941 ; MRC 1948; MRC 1950; MRC 1951). This 'random allocation' is the sole, but crucially important, feature of the category of fair tests referred to as 'randomized'. A random (as distinct from haphazard) allocation means that the chances of something happening are known, but the results cannot be anticipated on any particular occasion. So for example, if a coin is used to randomize, the chance of getting heads is 50%, but it is impossible to know what the result of a particular toss will be.

Casting or drawing lots is a time-honoured way of making fair decisions (Silverman and Chalmers 2002). These methods help to ensure that comparison groups are not composed of different types of people. Known and measured factors of importance, like age, can be checked. However unmeasured factors that may influence recovery from illness, such as diet, occupation, and anxiety, can be expected to balance out on average. If you would like to see how random allocation generates similar groups of people (click here for a demonstration).

As experience of using alternation and random allocation for unbiased assembly of groups of patients for comparing different treatments became more widespread, it became clear that strict adherence to unbiased allocation schedules was required to avoid biased creation of treatment comparison groups (MRC 1934). The risk of biased allocation can be abolished if treatment allocation schedules are concealed from those making decisions about participation in treatment comparisons – in brief, to prevent them cheating, and thus biasing the comparisons (MRC 1944; MRC 1948; MRC 1950; MRC 1951). The principle of random allocation to comparison groups can be applied both to individuals and to existing groups (for example, hospital wards, or general practices. The latter referred to as cluster randomization).

Avoiding biased losses from treatment comparison groups

After taking the trouble to ensure that treatment comparison groups are assembled in ways that ensure that like will be compared with like, it is important to avoid bias being introduced as a result of selective withdrawal of patients from the comparison groups. As far as possible, group similarity should be maintained by ensuring that all the people allocated to the treatment comparison groups are followed up and included in the main analysis of the test results – a so-called 'intention-to-treat' analysis (Bell 1941).

Failure to do this can result in unfair tests of treatments. Take, for example, two very different ways of treating people experiencing dizzy spells because of partially blocked blood vessels supplying their brains. Treatment for this condition can be important because people experiencing dizzy spells for this reason are at increased risk of suffering a stroke, which may leave them disabled, or even kill them. One of the treatments

for the dizzy spells involves taking aspirin to stop the blockage getting worse; the other involves a surgical operation to try to remove the blockage in the blood vessel.

A fair comparison of these two approaches to treating dizzy spells would involve creating two groups of people using an unbiased allocation method (like randomization), and then treating patients in one group with surgery and patients in the other group with aspirin. The comparison would thus begin by comparing two groups of patients who were alike, and go on to compare their respective frequencies of subsequent strokes. But if the frequency of strokes in the surgically treated group was only recorded among patients who had survived the immediate effects of the operation, the important fact that the operation itself can cause stroke and death would be missed. This would result in an unfair comparison of the two treatments, resulting in a biased and misleadingly optimistic picture of the effects of the operation. Like would not be being compared with like.

The principal comparison (in trials) must be based, as far as possible, on all the people assigned to receive each of the treatments compared, without exceptions, and in the groups to which they were originally assigned. If this principle is not observed, people may receive biased information about the overall effects of treatments.

< Previous Essay | Next Essay >

References

Balfour TG (1854). Quoted in West C. Lectures on the Diseases of Infancy and Childhood. London, Longman, Brown, Green and Longmans, p 600.

Bell JA (1941). Pertussis prophylaxis with two doses of alum-precipitated vaccine. Public Health Reports 56:1535-1546.

Cheselden W (1740). The anatomy of the human body. 5th edition. London: William Bowyer.

Colebrook D (1929). Irradiation and health. Medical Research Council Special Report Series No.131.

Doull JA, Hardy M, Clark JH, Herman NB (1931). The effect of irradiation with ultra-violet light on the frequency of attacks of upper respiratory disease (common colds). American Journal of Hygiene 13:460-77.

Lind J (1753). A treatise of the scurvy. In three parts. Containing an inquiry into the nature, causes and cure, of that disease. Together with a critical and chronological view of what has been published on the subject. Edinburgh: Printed by Sands, Murray and Cochran for A Kincaid and A Donaldson.

McPherson K (2004). Where are we now with hormone replacement therapy? BMJ 328:357-358.

Medical Research Council Therapeutic Trials Committee (1934). The serum treatment of lobar pneumonia. BMJ 1:241-245.

Medical Research Council (1944). Clinical trial of patulin in the common cold. Lancet 2:373-5.

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.

Medical Research Council (1951). The prevention of whooping-cough by vaccination. BMJ 1:1463-1471

Parry CH (1786). Experiments relative to the medical effects of Turkey Rhubarb, and of the English Rhubarbs,

No. I and No. II made on patients of the Pauper Charity. Letters and Papers of the Bath Society III:407-422.

Silverman WA, Chalmers I (2002). Casting and drawing lots: a time-honoured way of dealing with uncertainty and for ensuring fairness. JLL Bulletin: Commentaries on the history of treatment evaluation (http://jameslindlibrary.org/articles/casting-and-drawing-lots-a-time-honoured-way-of-dealing-with-uncertainty-and-for-ensuring-fairness/)

Theobald GW (1937). Effect of calcium and vitamin A and D on incidence of pregnancy toxaemia. Lancet 2:1397-1399.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

2.3 Why avoiding differences between treatments allocated and treatments received is important

Cite as: The James Lind Library 2.3 Why avoiding differences between treatments allocated and treatments received is important (http://jameslindlibrary.org/essays/2-3-why-avoiding-differences-between-treatments-allocated-and-treatments-received-is-important/)

> Knowledge of which treatments have been received by which study participants can affect adherence to assigned treatments and result in the biased use of other treatments (co-interventions). These biases can be reduced by using placebos to conceal the identities of the treatments being compared.

Fair tests of medical treatments have to be planned carefully. The documents setting out these plans are referred to as protocols, and, among other things, they specify details about the treatments that will be compared. The best laid plans don't always work out quite as intended, however. The treatments actually received by patients in tests sometimes differ from those it was intended they should have received. These departures from intention need to be taken into account in interpreting the results of treatment comparisons. One of the reasons that placebos were introduced in the evolution of fair tests of medical treatments was to reduce departures from intended treatments (Kaptchuk 1998).

Things may go astray even in placebo controlled trials, however. During the 2nd World War, people suffering from colds were given a solution of drug called patulin and compared with other people given only the fluid in which the drug had been dissolved (MRC 1944). Analysis of the results failed to reveal any beneficial effects of the drug, but then a concern emerged that the liquid used to dissolve the drug might have inactivated it. In other words, over 1000 patients might have participated in a comparison of two inactive treatments! Fortunately, tests confirmed that the patulin used in the trial had indeed been active, although it had no detectable effects on colds (Chalmers and Clarke 2004)!

Treatments received may differ from treatments intended for a variety of reasons. For example, doctors may decide that the treatment to which some of their patients have been allocated in a formal treatment comparison should not be offered to them; patients may reject the treatments allocated to them, or not take them as intended; doses of the treatment different from those intended may be given; or the supply of one of the treatments may run out.

For example, when differences emerged in the results of apparently identical treatments for leukaemia in British and American children, investigation revealed that the worse results in Britain reflected unwillingness among British clinicians to persist with chemotherapy when nasty toxic effects of treatment developed (Medical Research Council Working Party on Leukaemia in Children 1986).

For these reasons, interpretations of fair tests must consider the possibility that treatments received were not those intended, or that additional treatments were given to patients in one treatment comparison group to a greater extent than to those in another. If discrepancies between intention and practice have occurred, it is important to consider the possible implications for interpreting the evidence.

< Previous Essay | Next Essay >

References

Chalmers I, Clarke M (2004). The 1944 Patulin Trial: the first properly controlled multicentre trial conducted under the aegis of the British Medical Research Council. International Journal of Epidemiology 32:253-260.

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

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

Medical Research Council Working Party on Leukaemia in Children (1986). Improvement in treatment for children with acute lymphoblastic leukaemia. Lancet 1:408-11.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

2.4 The need to avoid differences in the way treatment outcomes are assessed

Cite as: The James Lind Library 2.4 The need to avoid differences in the way treatment outcomes are assessed (http://jameslindlibrary.org/essays/2-4-the-need-to-avoid-differences-in-the-way-treatment-outcomes-are-assessed/)

Biased treatment outcome assessment can result if people receiving or providing care, or others assessing treatment outcomes, know which participants have received which treatments. It is sometimes possible to conceal which treatments have been received by using placebos and in other ways.

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; Savovic et al. 2012). The greater the element of subjectivity in assessing outcomes, the greater the need to reduce these observer biases to ensure fair tests of treatments.


RAPPORT
DES COMMISSAIRES
CHARGÉS PAR LE ROI,
DE L'EXAMEN
DU
MAGNÉTISME ANIMAL.
Imprimé par ordre du Roi.
A PARIS,
DE L'IMPRIMERIE ROYALE.
M. DCC. LXXXIV.

In these common circumstances, 'blinding' of patients and doctors is a desirable element of fair tests. What appears to have been the earliest blinded (masked) assessment of a treatment was 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, one of the nineteenth century's major forms 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, anglophone researchers had taken up the use of 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). In the 1960s, 'Double dummies' were introduced when two very different treatments – an injection and a pill, for example – were being compared (Marušić A, Fatović-Ferenčić S 2012).

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 with 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.

< Previous Essay | Next Essay >

References

Bingel A (1918). Uber Behandlung der Diphtherie mit gewohnlichem Pferdeserum. Deutsches Archiv fur 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). Receuil 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.

Lohner G (1835), on behalf of a Society of truth-loving men. Die Homooopathischen Kochsalzversuche zu Nurnberg [The homeopathic salt trials in Nuremberg].

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

Marušić A, Fatović-Ferenčić S (2012). Adoption of the double dummy trial design to reduce observer bias in testing treatments

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.

Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, Als-Nielsen B, Balk EM, Gluud C, Gluud LL, Ioannidis JPA, Schulz KF, Beynon R, Welton NJ, Wood L, Moher D, Deeks JJ, Sterne JAC (2012). Influence of reported study design characteristics on intervention effect estimates from randomized controlled trials. Annals of Internal Medicine 157:429-438.

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.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

2.5 Bias introduced after looking at study results

Cite as: The James Lind Library 2.5 Bias introduced after looking at study results (http://jameslindlibrary.org/essays/2-5-bias-introduced-after-looking-at-study-results/)

> Biases can be introduced when knowledge of the results of studies influences analysis and reporting decisions, for example, when studies stop earlier than planned, or if there is bias in the selection of the treatment outcomes analyzed.

Bias results from processes that tend to produce information that depart systematically from the truth. Avoiding bias is relevant when analysing the results of studies statistically. Analysis biases may be introduced during the design of studies, when decisions about which analyses to do might lead to the favouring of one of the treatments compared over another. This might include decisions about how to deal with data for participants who don't adhere to their allocated intervention, the analysis of those who experience other outcomes before the main outcomes for the study, or the definition, counting and combination of particular outcomes in the analyses. These design biases are akin to those that can arise when the choice of the comparator to test in the study has been biased so that the eventual results will be unduly favourable to the newer treatment.

Things can get much worse after study results have been inspected. Changes might then be made to how the analyses will be done or reported, with fore-knowledge of how these changes will favour one or other of the treatments compared. If these changes occur between the collection of the study data and its eventual reporting, the reader of the published results might be misled, especially if the changes are not clearly described and explained.

Biased analyses before the planned end of a study

Biases after looking at study results can occur both after formal statistical analyses, and through more informal routes. For example, if the researchers are collecting the outcomes or observing these outcomes because they are providing the treatments to participants in the study, they may get a sense of the accumulating results, for example, about which patients are doing particularly well or badly. This might lead them to alter the planned analyses, such as changing what they feel is the "most important" outcome, choosing an earlier time point as the main one to emphasise, or dividing the data in different ways in subgroup analyses. One way to avoid this is by keeping the researchers and the practitioners blind (masked) to the treatment allocated to each participant.

When study results are being analysed more formally, the problems can become worse as these initial analyses may reveal what the results would be if the analyses are modified. Such biases might occur before or after the study has reached its intended completion.

During a study, accumulating results might be examined to see if there is clear evidence of benefit or harm for one intervention, which might make it unethical to continue the study. On the other hand, it may become clear that the effect that was hoped for is not achievable in the study and that it would be better to stop the study for futility rather than to continue to recruit participants to a study that will use resources but will not resolve the initial uncertainties. These early stopping decisions can lead to bias when the interim results happen to be high or low simply by chance, especially if there is a vested interest in closing the study and turning these interim results into its final results (Trotta 2008).

One way to avoid biases that might arise if the researchers themselves are responsible for these interim decisions is to have an independent Data Monitoring Committee consider the accumulating results. The committee can agree guidelines for deciding when to make interim analyses available to an oversight group for the study, such as a Trial Steering Committee (Grant 2005).

Sometimes, interim results may be presented more publicly, to allow practitioners and potential participants in the trial to make up their own minds about whether or not to continue with the study. For example, the preliminary results of the ISIS-2 trial of aspirin and streptokinase for people having a heart attack (myocardial infarction). The trial Steering Committee published a half-page interim report showing benefits reported to them the previous month. These showed a reduction in the risk of death in the short term among patients who had received streptokinase within 4 hours of experiencing symptoms of heart attack (ISIS-2 1987). Despite this information, some insufficiently persuaded clinicians continued to recruit patients to the trial within this time window, as well as others who had presented more that 4 hours after their symptoms had begun (ISIS-2 1988).

Biased analyses after the planned end of a study

At the end of a study, changes to the analyses after looking at the results can lead to bias through:

- changes in the primary outcome, or in how outcomes are defined or combined in composite outcomes;
- introduction or modification of subgroup analyses, in which different groups of participants are analysed separately; perhaps to highlight the presence or absence of benefit in certain types of person or setting. In addition to the problems of bias in these analyses, chance might mean that the findings are not a reliable guide to the truth (Counsell 1994, Clarke 2001);
- selective reporting of particular outcomes, analyses or treatment comparisons. For example, in a study comparing three treatments, there are seven different ways in which the treatments might be compared. This gives researchers opportunities to highlight some comparisons over others, based purely on their results; and
- changes to the statistical techniques, such as the introduction of adjustments for differences in the baseline characteristics of the participants which had not been pre-planned or pre-specified.

The potential impact of some of these biases have been studied, and some of these studies have themselves have been considered in systematic reviews. For example, systematic reviews by Kerry Dwan and colleagues have brought together information on how the methods used in the analyses and reporting of randomised trials changed between the design phase of the trial and the publication of its results.

In their most recent review, they found 22 studies (containing more 3000 randomised participants) published between 2000 and 2013 that found discrepancies in statistical analyses (8 studies), composite outcomes (1), the handling of missing data (3), unadjusted versus adjusted analyses (3), handling of continuous data (3) and subgroup analyses (12), concluding that discrepancies in analyses between publications and other study documentation were common, but not discussed in the trial reports (Dwan 2014). In their systematic reviews of studies of selective reporting, they found that comparisons of trial publications to protocols found that 40-62% of studies had at least one primary outcome that was changed, introduced, or omitted (Dwan 2011; Dwan 2013).

In systematic reviews of the impact of early stopping, Montori et al in 2005 and Bassler et in 2010 have shown how early stopping might bias conclusions about the effects of interventions. The first review included 143 randomised trials stopped early for benefit, with 92 of these published in 5 high-impact, influential medical journals and, on average, the trials recruited about two-thirds of their planned sample size. Montori et al concluded that randomised trials stopped early for benefit were becoming more common, often fail to adequately report relevant information about the decision to stop early, and show implausibly large treatment effects, particularly when the number of events is small. They wrote that “clinicians should view

the results of such trials with scepticism" (Montori 2005). Five years later, Bassler et al compared 91 truncated randomised trials with 424 matched non-truncated trials, finding a pooled ratio of relative risks of 0.71 (95% confidence interval, 0.65-0.77). This showed that the effects estimates in the trials that stopped early were on average more favourable to the treatments than those from similar trials that did not stop early (Bassler 2010).

If users of the reports of studies are to have confidence in their final reports, they need to be reassured that bias was not introduced to the results in those reports after the early results had been seen. Although the afore-mentioned reviews show that protocols are no guarantee against this, access to a protocol or a study's statistical analysis plan might identify any changes that were made; and, since 2013, guidance on the structured reporting of protocols is available from the SPIRIT group (Chan 2013). In relation to the choice of outcomes to analyze and report, those designing studies should consider the use of core outcome sets as the minimum that they should measure, analyze and report in all trials in a particular condition. Work by the COMET initiative has already identified 200 such outcome sets (Gargon 2014), which are now available through the COMET database www.cometinitiative.org/studies/search.

It is tempting for people to change their views on what is important about a study after they have knowledge of the results. Such biases need to be avoided by careful planning of what analyses will be done, and clear explanations of any changes that were made to those plans and the reasons for them.

< Previous Essay | Next Essay >

References

Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, Heels-Ansdell D, Walter SD, Guyatt GH; STOPIT-2 Study Group (2010). Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. JAMA 303:1180-1187.

Chan A-W, Tetzlaff JM, Altman DG, Laupacis A, Gotzsche PC, Krleža-Jerić K, Hróbjartsson A, Mann H, Dickersin K, Berlin J, Doré C, Parulekar W, Summerskill W, Groves T, Schulz K, Sox H, Rockhold FW, Rennie D, Moher D (2013). SPIRIT 2013 Statement: Defining standard protocol items for clinical trials. Annals of Internal Medicine 158:200-207.

Counsell CE, Clarke MJ, Slattery J, Sandercock PAG (1994). The miracle of DICE therapy for acute stroke: fact or fictional product of subgroup analysis? BMJ 309:1677-1681.

Clarke M, Halsey J (2001). DICE 2: a further investigation of the effects of chance in life, death and subgroup analyses. International Journal of Clinical Practice 55:240-242.

Dwan K, Altman DG, Clarke M, Gamble C, Higgins JP, Sterne JA, Williamson PR, Kirkham JJ (2014). Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. PLoS Medicine 11(6):e1001666.

Dwan K, Altman DG, Cresswell L, Blundell M, Gamble CL, Williamson PR (2011). Comparison of protocols and registry entries to published reports for randomised controlled trials. Cochrane Database of Systematic Reviews (1):MR000031.

Dwan K, Gamble C, Williamson PR, Kirkham JJ; Reporting Bias Group (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias – an updated review. PLoS One 8(7):e66844.

Gargon E, Gurung B, Medley N, Altman DG, Blazeby JM, Clarke M, Williamson PR (2014). Choosing important health outcomes for comparative effectiveness research: a systematic review. PLoS ONE 9(6):e99111.

Grant AM, Altman DG, Babiker AB, Campbell MK, Clemens FJ, Darbyshire JH, Elbourne DR, McLeer SK, Parmar MK, Pocock SJ, Spiegelhalter DJ, Sydes MR, Walker AE, Wallace SA; DAMOCLES study group (2005). Issues in data monitoring and interim analysis of trials. Health Technology Assessment 9(7):1-238.

ISIS-2 Steering Committee (1987). Intravenous streptokinase given within 0-4 hours of onset of myocardial infarction reduced mortality in ISIS-2. Lancet 329:502.

ISIS-2 (second International Study of Infarct Survival) Collaborative Group (1988). Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17 187 cases of suspected acute myocardial infarction: ISIS-2. Lancet 332:349–360.

Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, Lacchetti C, Leung TW, Darling E, Bryant DM, Bucher HC, Schunemann HJ, Meade MO, Cook DJ, Erwin PJ, Sood A, Sood R, Lo B, Thompson CA, Zhou Q, Mills E, Guyatt GH (2005). Randomized trials stopped early for benefit: a systematic review. JAMA 294:2203-2209.

Trotta F, Apolone G, Garattini S, Tafuri G (2008). Stopping a trial early in oncology: for patients or for industry? Annals of Oncology 19(7):1347-1353

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

2.6 Reducing biases in judging unanticipated possible treatment effects

Cite as: The James Lind Library 2.6 Reducing biases in judging unanticipated possible treatment effects (http://jameslindlibrary.org/essays/2-6-reducing-biases-in-judging-unanticipated-possible-treatment-effects/)

> Important unanticipated effects of treatments are often first suspected by people using or prescribing treatments. As with anticipated effects of treatments, steps must be taken to reduce biases and the play of chance in assessing suspected unanticipated effects.

It is only to be expected that unanticipated effects of treatments will emerge when new treatments are introduced more widely. Initial tests – for example, those required to license new drugs for marketing – cover at most a few hundred or a few thousand people treated for a few months. Only relatively frequent and short-term unanticipated effects are likely to be picked up at this stage.

Rare treatment effects, or those that take some time to develop, will not be discovered until treatment tests have lasted long enough or until there has been more widespread use of treatments. Moreover, new treatments will often be used in people who may differ in important ways from those who participated in the original tests. They may be older or younger, of a different sex, more or less ill, living in different circumstances, or suffering from other health problems in addition to the condition at which the treatment is targeted. These differences may modify treatment effects, and new, unanticipated effects may emerge (see special issue of BMJ 3 July 2004).

Detection and verification of unanticipated effects, whether adverse (or beneficial), usually occur rather differently from the methods used to assess hoped-for effects of new treatments. Unanticipated effects of treatments are sometimes suspected initially by health professionals or patients. (Venning 1982) Identifying which among these initial hunches reflect real effects of treatments poses a challenge.

If the unanticipated effect of a treatment is very striking and occurs quite often after the treatment has been used, it may be noticed spontaneously by health professionals or patients. For example, babies born without limbs are almost unheard of, so when a sudden increase in their numbers occurred in the 1960s it naturally raised concerns. All mothers of such babies had used a newly marketed anti-nausea drug – thalidomide – prescribed during early pregnancy, so this was likely to be the cause and little further assessment was necessary (McBride 1961). Unanticipated beneficial effects of drugs are often detected in similar ways, for example, when it was found that a drug to treat psychosis also lowered cholesterol (Goodwin 1991).

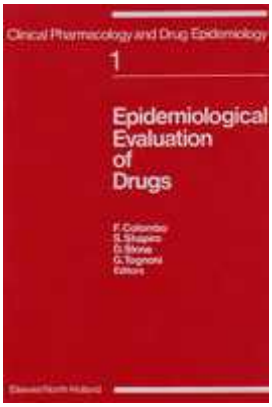
When such striking relationships are noticed, they often turn out to be confirmed as real unanticipated effects of treatment (Venning 1982). However, a lot of hunches about unanticipated effects of treatment are based on far less convincing evidence. So, as with tests designed to detect hoped-for effects of treatments, planning tests to confirm or dismiss less striking suspected unanticipated effects involves avoiding biased comparisons [EE 2.0]. Studies to test whether suspected unanticipated effects of treatment are real must observe the principle of comparing 'like with like'. Random allocation to treatments is the ideal way to accomplish this. Only rarely, however, can suspected treatment effects be investigated by further analysis or follow-up of people who were randomly allocated to treatments before they were given (Hemminki and

McPherson 1997). The challenge is therefore to assemble unbiased comparison groups in other ways, often using information collected routinely during health care.

In these studies, it actually helps that the suspect effects were not anticipated at the time that treatment decisions were taken. This is because it means that no account could have been taken of the risk of the suspect condition at the time people were being selected differentially for treatment: the unanticipated effect is usually a different condition or disease from the condition or disease for which the treatment was prescribed (Vandenbroucke 2004a).

For example, when hormone replacement therapy (HRT) was introduced for treating menopausal symptoms, a woman's risk of developing venous thrombosis was unlikely to have been taken into account because most doctors and women thought it was irrelevant. There was therefore no reason to expect that women who were prescribed HRT differed in their risk of developing venous thrombosis from those who did not receive the drug. The basis was thus established for fair tests, and these showed that HRT increases the risk of venous thrombosis.

When a suspected unanticipated effect relates to a treatment for a common health problem (such as heart attack) but does not occur very often with the new treatment (or is not completely relieved by it), large-scale surveillance of people receiving the treatment is needed to detect the unanticipated effect. For example, although some people thought that aspirin might reduce the risk of heart attack and began fair tests of this theory in patients in the late 1960s (Elwood et al. 1974), most people would have thought that the theory was highly implausible. The breakthrough came when a large study was done to detect unanticipated adverse effects of drugs: researchers noticed that people admitted to hospital with heart attacks were less likely to have recently taken aspirin than apparently similar patients (Boston Collaborative Drug Surveillance Group 1974). These findings were consistent with those of a fair test, in which people had been allocated at random to receive or not receive aspirin after heart attack. The two reports were published back-to-back in the same issue of the British Medical Journal .



The ground rules for detecting and investigating unanticipated effects of treatments were first set out clearly in the late 1970s (Jick 1977; Colombo et al. 1977). They drew on the collective experience of investigating unanticipated effects which had accumulated following the thalidomide disaster. The requirements for one important type of research, case-control studies of possible adverse effects of treatment, were laid down in a paper based on the experiences of researchers in Boston and Oxford (Jick and Vessey 1978). With many powerful treatments introduced since that time, this aspect of fair tests of treatments remains just as challenging and important today as it did then (Vandenbroucke 2004b; Vandenbroucke 2006; Papanikolaou et al. 2006).

It is important to recognise that individual reports suggesting or dismissing suspicions about unanticipated effects of treatments can be misleading. As with all other fair tests of treatment, possible unanticipated effects of treatment must be investigated using systematic reviews of all the relevant evidence, such as those that confirmed the relationship between HRT and heart disease, stroke and breast cancer (Hemminki and McPherson 1997; Collaborative Group on Hormonal Factors in Breast Cancer 1997).

< Previous Essay | Next Essay >

References

Boston Collaborative Drug Surveillance Group (1974). Regular aspirin intake and acute myocardial infarction. BMJ 1:440-443.

Collaborative Group on Hormonal Factors in Breast Cancer (1997). Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Lancet 350:1047-1059 .

Colombo F, Shapiro S, Slone D, Tognoni G, eds (1977). Epidemiological Evaluation of Drugs. Amsterdam: Elsevier/North Holland Biomedical Press, 1977.

Elwood PC, Cochrane AL, Burr ML, Sweetnam PM, Williams G, Welsby E, Hughes SJ, Renton R (1974). A randomised controlled trial of acetyl salicylic acid in the secondary prevention of mortality from myocardial infarction. BMJ 1:436-440.

Goodwin JS (1991). The empirical basis for the discovery of new therapies. Perspectives in Biology and Medicine 35:20-36.

Hemminki E, McPherson K (1997). Impact of postmenopausal hormone therapy on cardiovascular events and cancer: pooled data from clinical trials. BMJ;315:149-153.

Jick H (1977). The discovery of drug-induced illness. New England Journal of Medicine 296:481-485.

Jick H, Vessey M (1978). Case-control studies in the evaluation of drug-induced illness. American Journal of Epidemiology 107:1-7.

McBride WG (1961). Thalidomide and congenital abnormalities. Lancet 2:1358.

Papanikolaou PN, Christidi GD, Ioannidis JPA (2006). Comparison of evidence on harms of medical interventions in randomized and nonrandomsized studies. CMAJ 174:635-641.

Vandenbroucke JP (2004a). When are observational studies as credible as randomised trials? Lancet 363:1728-1731.

Vandenbroucke JP (2004b). Benefits and harms of drug treatments. BMJ 329:2-3.

Vandenbroucke JP (2006). What is the best evidence for determining harms of medical treatment? CMAJ 174:645-646.

Venning GR (1982).Â Validity of anecdotal reports of suspected adverse drug reactions: the problem of false alarms.Â BMJ 284:249-254.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

2.7 Dealing with biased reporting of the available evidence

Cite as: The James Lind Library 2.7 Dealing with biased reporting of the available evidence (http://jameslindlibrary.org/essays/2-7-dealing-with-biased-reporting-of-the-available-evidence/)

> Biased reporting of research occurs when the direction or statistical significance of results influences whether and how research is reported

Avoiding biased comparisons entails identifying and taking account of all the relevant reliable evidence in systematic reviews. This is challenging in many ways, particularly as some pertinent evidence is not published because biased decisions are made about which results of research are submitted and accepted for publication. Studies that have yielded 'disappointing' or 'negative' results are less likely to be reported than others. This is often called 'publication bias' or 'reporting bias'. It might arise from biased analyses of studies, after their results are known.

These reporting biases have been recognized for centuries (Dickersin and Chalmers 2010). In 1792, for example, James Ferriar stressed the importance of recording treatment failures as well as treatment successes (Ferriar 1792). This principle was reiterated in an editorial published in the Boston Medical and Surgical Journal just over a century later (Editorial 1909).

There is now a large body of evidence confirming that reporting bias is a substantial problem. There is also evidence that reporting bias results principally from researchers not writing up or submitting reports of research for publication, not because of biased rejection of submitted reports by journal editors (Dickersin 2004). Recent research has also revealed an additional problem: if estimates of treatment effects on some of the outcomes studied don't support the conclusions of researchers, these data sometimes don't get reported either (Chan et al. 2004).

For example, had all the studies of the effects of giving drugs to reduce heart rhythm abnormalities in patients having heart attacks been reported, tens of thousands of deaths from these drugs could have been avoided. In 1993, Dr Cowley and his colleagues pointed out how an unpublished study done 13 years previously might have "provided an early warning of trouble ahead". Nine patients had died among the 49 assigned to the anti-arrhythmic drug (lorcainide) compared with only one patient among a similar number given placebos. "When we carried out our study in 1980", they reported, "we thought that the increased death rate was an effect of chance…The development of lorcainide was abandoned for commercial reasons, and this study was therefore never published; it is now a good example of 'publication bias'" (Cowley et al. 1993).

Reporting biases tend to lead to conclusions that medical treatments are more useful and freer of side effects than they are in fact. They can therefore result in unnecessary suffering and death, and in wasted resources spent on ineffective or dangerous treatments (Chalmers 2004). People who agree to researchers' requests that they participate in tests of treatments assume that their participation will lead to an increase in knowledge. This implied contract between researchers and participants in research is breached by researchers who do not make public the results of the research.

Biased under-reporting of research is scientific misconduct and unethical (Chalmers 1990). Selective reporting of studies sponsored by the pharmaceutical industry is a particular problem (Hemminki E 1980; Melander et al. 2003), although the problem is not limited to those with commercial vested interests. Research ethics committees, medical ethicists and research funders have so far not done enough to protect patients and the public from the adverse effects of reporting biases (Savulescu et al. 1996). Fair testing of treatments – particularly those treatments in which there is commercial interest – will remain compromised just as long as this form of research misconduct is tolerated by governments and others who should be protecting the interests of the public.

The World Health Organization has begun to coordinate solutions to address the problem of unidentifiable research and publication (or dissemination) bias: First, it has established standards for the registration and exchange of data for the registration of trials. Secondly, it proposes registration of research protocols in databases that fulfill the above standards, before patient recruitment starts. Finally, it has established an open access portal (www.who.int/ictrp), which collates the data of all national and regional registers, allowing people to learn about coming, ongoing and finished research protocols. Since 2013, the All Trials Campaign (www.alltrials.net) has called for registration and reporting of all trials.

< Previous Essay | Next Essay >

References

Chalmers I (1990). Under-reporting research is scientific misconduct. JAMA 263:1405-1408.

Chalmers I (2004). In the dark: drug companies should be forced to publish all the results of clinical trials. New Scientist 181:19.

Chan A-W, Hròbjartsson A, Haahr M, Gøtzsche PC, Altman DG (2004). Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to publications. JAMA 291:2457-2465.

Cowley AJ, Skene A, Stainer, Hampton JR (1993). The effect of lorcainide on arrhythmias and survival in patients with acute myocardial infarction. International Journal of Cardiology 40:161-166.

Dickersin K (2004). How important is publication bias? A synthesis of available data. AIDS Educ Prev 1997;9 (1 Suppl):15-21.

Dickersin K, Chalmers I (2010). Recognising, investigating and dealing with incomplete and biased reporting of clinical research: from Francis Bacon to the World Health Organisation. JLL Bulletin: Commentaries on the history of treatment evaluation, (www.jameslindlibrary.org)

Editorial (1909). The reporting of unsuccessful cases. Boston Medical and Surgical Journal 161:263-264.

Ferriar J (1792). Medical histories and reflexions. Vol 1. London: Cadell and Davies, 1792.

Hemminki E (1980). Study of information submitted by drug companies to licensing authorities. BMJ 280:833-6.

Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B (2003). Evidence b(i)ased medicine – selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. BMJ 326:1171-3.

Savulescu J, Chalmers I, Blunt J (1996). Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. BMJ 313:1390-1393.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

2.8 Avoiding biased selection from the available evidence

Cite as: The James Lind Library 2.8 Avoiding biased selection from the available evidence (http://jameslindlibrary.org/essays/2-8-avoiding-biased-selection-from-the-available-evidence/)

> Because single tests of treatments can be misleading, systematic reviews are used to identify, evaluate and summarize all the evidence relevant to addressing a particular question.

Biases can distort individual tests of medical treatments and lead to erroneous conclusions. They can also distort reviews of evidence. Plans for systematic reviews should be set out in protocols, such as those published by The Cochrane Collaboration, making clear what measures will be taken to reduce biases.

These include specifying clearly:

- which questions about treatments will be addressed in the review;
- the criteria that will make a study eligible for inclusion;
- the strategies that will be used to search for potentially eligible studies; and
- the steps that will be taken to minimise biases in selecting studies and data for inclusion in the review (Berlin 1997).

Different systematic reviews addressing what appears to be the same question about the effects of medical treatments quite often reach different conclusions. Sometimes this is because the questions addressed are subtly different. Sometimes it reflects differences in the materials and methods used by the reviewers, and in these circumstances it is important to judge which of the reviews are most likely to have reduced biases most successfully.

It is also worth considering whether the reviewers have other interests that might affect the conduct or interpretation of their review. For example, people associated with the manufacturers of evening primrose oil reviewed the drug's effects on eczema (Morse et al. 1989). They reached a far more enthusiastic conclusion about the value of the drug than a review done by investigators with no commercial interest, who included the results of unpublished studies in their assessment (Williams 2003).

It is not only commercial interests that can lead to biased selection from the available evidence for inclusion in reviews. We all have prejudices that can lead to biased selection of evidence, and researchers, health professionals, patients and others assessing the effects of treatments are not immune.

< Previous Essay | Next Essay >

References

Berlin JA (1997). Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Blinding Study Group. Lancet 350:185-186.

Morse PF, Horrobin DF, Manku MS, Stewart JC, Allen R, Littlewood S, Wright S, Burton J, Gould DJ, Holt PJ, et

al (1989). Meta-analysis of placebo-controlled studies of the efficacy of Epogam in the treatment of atopic eczema. Relationship between plasma essential fatty acid changes and clinical response. British Journal of Dermatology 121:75-90.

Williams HC (2003). Evening primrose oil for atopic dermatitis. BMJ 327:1358-1359.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

2.9 Recognizing researcher/sponsor biases and fraud

Cite as: The James Lind Library 2.9 Recognizing researcher/sponsor biases and fraud (http://jameslindlibrary.org/essays/2-9-recognizing-researchersponsor-biases-and-fraud/)

> The commercial, academic or other vested interests of researchers and organizations tend to be reflected in reports of treatment research in which they are involved.

In 1764, a Dr R James published the 6th edition of his book 'A dissertation on fevers and inflammatory distempers'. In it, he claimed that his secret 'Fever Powder' was successful in treating "smallpox, yellow fever, slow fever and rheumatism". In support of his claims, he cited the testimonies of satisfied patients and a decline in the national mortality rate which had followed his introduction of his miraculous 'cure-all'. 'Snake oil salesmen' like Dr James have probably been a feature of medical practice for as long as patients have looked to doctors and others to help them deal with health problems.

A few years after Dr James made his claims, Elisha Perkins patented and marketed 'tractors' – small metal rods which he claimed cured a variety of ailments through 'electrophysical force'. The cartoon accompanying this paragraph shows a doctor treating a wealthy client with Perkins' tractors. John Haygarth (1800) conducte d an experiment to test the claims made by another salesman. In a pamphlet entitled 'Of the imagination as a cause and as a cure of disorders of the body: exemplified by fictitious tractors', Haygarth reported how he put Perkins' claims to the 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).

During the 19th century, the ground rules for testing treatment claims began to become clearer. Alternation began to be used to ensure that like would be compared with like (Chalmers et al. 2011); and blinding became recognised as a way of reducing observer biases (Kaptchuk 2011). For example, comparisons of homeopathic with orthodox medical treatments (Löhner 1835; Tessier 1852; Storke et al. 1880; Cook County Board of Commissioners 1882) demonstrated not that homeopathy was effective, but that it was safer than the bleeding and purging being offered by mainstream doctors.

By the early years of the 20th century, a pharmaceutical industry had begun to emerge which was profit-driven, and thus tempted to take liberties in its claims for its products and the use of data to support these. In 1917, Torald Sollmann, an American pharmacologist, set out the principles to be observed in testing treatments, and noted that "Those who collaborate with [commercial firms] should realize frankly that under present conditions they are collaborating, not so much in determining scientific value, but rather in establishing commercial value" (Sollmann 1917). Concerns about these sponsor and researcher biases – and sometimes frank fraud – grew throughout the 20th century, fuelled increasingly by evidence going beyond anecdotes (Davidson 1986; Gøtzsche 1987; Djulbegovic et al. 2000; Lexchin et al. 2003; Melander et al. 2003; Whittington et al 2004; Yank et al. 2007; Turner et al. 2008; Rising et al. 2008). Sponsor and researcher biases make active use of other biases in pursuit of their vested interests, particularly design bias, analysis bias, and

reporting bias.

Recognising and reducing research biases and frank fraud remains a substantial challenge.

< Previous Essay | Next Essay >

References

Chalmers I, Dukan E, Podolsky SH, Davey Smith G (2011). The advent of fair treatment allocation schedules in clinical trials during the 19th and early 20th centuries. JLL Bulletin: Commentaries on the history of treatment evaluation (www.jameslindlibrary.org).

Cook County Board of Commissioners (1882). Proceedings, 1881-1882. Chicago, Illinois, 222-223.

Davidson RA (1986). Source of funding and outcome of clinical trials. Journal of General Internal Medicine 1:156-158.

Djulbegovic B, Lacevic M, Cantor A, Fields KK, Bennett CL, Adams JR, Kuderer NM, Lyman GH (2000). The uncertainty principle and industry-sponsored research. Lancet 356:635-638.

Gøtzsche PC (1987). Reference bias in reports of drug trials. BMJ 295:654-656.

James R (1764). A dissertation on fevers and inflammatory distempers. 6th edn. London: Newbery.

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 (2011). A brief history of the evolution of methods to control of observer biases in tests of treatments. JLL Bulletin: Commentaries on the history of treatment evaluation (www.jameslindlibrary.org).

Lexchin J, Bero LA, Djulbegovic B, Clark O (2003). Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ 326:1167-70.

Löhner G (1835). Die Homoöopathischen Kochsalzversuche zu Nürnberg [The homeopathic salt trials in Nurnberg]. Nürnberg in März.

Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B (2003). Evidence b(i)ased medicine – selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. BMJ 326:1171-3.

Rising K, Bacchetti P, Bero L (2008). Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. PLoS Med 5(11): e217. doi:10.1371/journal.pmed.0050217

Sollmann T (1917). The crucial test of therapeutic evidence. JAMA 69:198-199.

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. Homoeopathic Times 7:12/280-1.

Tessier JP (1852). De la médication homoeopathique [On homoeopathic medication]. Paris: JB Ballière.

Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. New England Journal of Medicine 358:252-60.

Whittington CJ, Kendall T, Fonagy P, Cottrell D, Cotgrove A, Boddington E (2004). Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. Lancet; 363:1341-1345.

Yank V, Rennie D, Bero LA (2007). Financial ties and concordance between results: retrospective cohort study.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

3.0 Taking account of the play of chance

Cite as: The James Lind Library 3.0 Taking account of the play of chance (http://jameslindlibrary.org/essays/3-0-taking-account-of-the-play-of-chance/)

> When treatments are compared, any differences in outcome events may simply reflect the play of chance. Increasing the number of events studied in research reduces the likelihood of being misled by the play of chance.

When two treatments are compared, any differences in outcome may simply be caused by the play of chance. For example, take a comparison of a new treatment with a standard treatment in which 4 people improved with the former and 6 people improved with the latter. It would clearly be wrong to conclude confidently that the new treatment was worse than the standard treatment: these results might simply reflect the play of chance. If the comparison was repeated, the numbers of patients who improved might be reversed (6 against 4), or come out the same (5 against 5), or in some other ratio.

If, however, 40 people improved with the new treatment and 60 with the standard treatment, chance becomes a less likely explanation for the difference. And if 400 people improved with the new treatment and 600 with the standard treatment, it would be clear that the new treatment was indeed very likely to be worse than the standard. The way to reduce the likelihood of being misled by the play of chance in treatment comparisons is thus to ensure that fair tests include sufficiently large numbers of people who experience the outcomes one hopes to influence, such as increased improvement or reduced deterioration.

In some circumstances very large numbers of people – thousands and sometimes tens of thousands – need to participate in fair tests to obtain reliable estimates of treatment effects. Large numbers of participants are necessary, for example, if the treatment outcomes of interest are rare – for example, heart attacks and strokes among apparently healthy middle-aged women using hormone replacement therapy (HRT). Large numbers are also needed if moderate but important effects of treatments are to be detected reliably – for example, a reduction by 20 per cent in the risk of early death among people having heart attacks.

< Previous Essay | Next Essay >

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

3.1 Recording and interpreting numbers in testing treatments

Cite as: The James Lind Library 3.1 Recording and interpreting numbers in testing treatments (http://jameslindlibrary.org/essays/3-1-recording-and-interpreting-numbers-in-testing-treatments/)

Numbers are needed to record the results of fair tests of treatments, and tables and graphs are used to describe the characteristics and experience of groups of patients, the treatment they have received, and quantitative estimates of treatment effects.

Using quantification in testing treatments

It was not until the early 18th century that numbers began to be used to assess the effects of medical interventions (Nettleton 1722a; 1722b; 1722c; Jurin 1724; and Huth 2005; Boylston 2010; Boylston 2012). This development occurred chiefly, but not exclusively, in Britain, Where it has been systematically studied (Tröhler 2010) . In 1732, for example, Francis Clifton published a book entitled The state of physick, ancient and modern, briefly considered: with a plan for the improvement of it (Clifton 1732). Clifton pointed out that, instead of assessing the worth of therapies by whether they accorded with theories, physicians needed to base their judgements about the effects of treatments on a sufficient number of their own (or otherwise testified) observations, organised in tables. A series of authors emphasised similar principles throughout the 18th century. Quantification began in the 1720s with comparisons of death rates following variolation (inoculation) with those associated with the disease itself, establishing the relative safety of the former. In this field, numbers were used throughout the century in various European countries, and they played a role in the introduction of vaccination around 1800. By then numerical data had become an important criterion for assessing new therapies in surgery and medicine (Tröhler 2010).

Using tables and graphs to present treatment comparisons

In the British Army, John Rollo (1781) may have been the first to use tables to give a detailed account of all the cases he had treated in a military hospital in Barbados. When he became chief of the Hospital of the Ordnance (artillery) at Woolwich, he published a hospital report based on the same principles (Rollo 1801). Richard McCausland (1783) published his comparative studies of various treatments for intermittent fevers in statistical tables (Maehle 2011). Thomas Dickson Reide's (1793) View of the diseases of the Army was full of tabular compilations and arithmetical calculations, as were James McGrigor's reports (1801; 1815). Numerical data were quite often used to calculate simple ratios. For example, William Falconer (1807) calculated success:failure ratios to compare the results of his practice in Bath with those published earlier by Rice Charleton (1770).

Replacing certainties with probabilities

What were the motives for quantifying and tabulating oberservations? What were the numbers intended to convey? The title and the place of publication of a work by George Fordyce provide an initial answer: An attempt to improve the evidence of medicine (Fordyce 1793), published in the Transactions of a Society for the Improvement of Medical and Chirurgical Knowledge. Quantification of experience aimed at "increasing the certainty of medicine." John Millar (1798) observed that "Where mathematical reasoning can be had, it is a great folly to make use of any other, as to grope for a thing in the dark, when you have a candle standing by you"; and his dispensary-physician colleague William Black (1789) noted that: "However it may be slighted as an heretical innovation, I would strenuously recommend Medical Arithmetick as a guide and compass through the labyrinth of therapeutick." Reide (1793) justified this approach using a simple analogy:" How

ridiculous would it appear [for a merchant] to judge of the advantages or disadvantages of particular branches of commerce from reasoning and conjecture whilst the result can be reduced to certainty be keeping regular accounts, and balancing them at stated periods."

Methodological questions were indeed eagerly debated in 18th century British medicine. Among the issues there was that of certainty versus the slowly growing notion of statistical probability. In 1772, for example, James Lind, then chief of the 1000-bed Haslar Navel Hospital, summarized the transition from belief in an absolute authority to reliance on relative statistics: "A work indeed more perfect, and remedies more absolutely certain might perhaps have been expected from an inspection of several thousand…patients." But even such facts always remained partial in his view, and he concluded with the remarkable insight that "for though they may for a little, flatter with hopes of greater success, yet more enlarged experience must ever evince the fallacy of all positive assertions in the healing art" (Lind 1772, p v-vi).

More outspokenly, John Haygarth, with the help "of an ingenious friend, Mr Dawson, a truly mathematical genius", calculated probabilities of escaping infection by 'continuous fever' or smallpox. On the basis of results "computed arithmetically by the doctrine of chances, according to the data", he advocated the immediate isolation of smallpox and fever patients in specific wards in Chester (Haygarth 1784, p 26-28).

Interpreting numbers

With the availability of more and more numerical data, numbers began to be pitted against numbers at the beginning of the 19th century. How did people judge whether treatment comparisons were trustworthy and meaningful? For example, during the debates about bloodletting for the treatment of fevers around 1800, statistics were widely used on both sides. It became clear that these data needed interpretation. In 1813, Thomas Mills had re-introduced copious bloodletting and purging at the Dublin Fever Hospital. The statistics comparing his mortality rates with those of other physicians who had hardly used bloodletting were reprinted in the review of his Essay on the utility of blood-letting in fever (Mills 1813). This elicited the following comment:

> presuming….these are candid and correct statements, we may deem them potent arguments in favour of the advantages of the anti-phlogistic [bloodletting and purging] treatment of fever (Edinburgh Medical and Surgical Journal 1813).

Besides the issue of honesty the question of bias was raised, of the need to compare the comparable. For instance, the Monthly Review wrote that Mills' work left "a rather painful impression on our minds," for these impressive results might be explained by the type of patients treated by Mills rather than by the therapy he had applied (Monthly Review 1814, p314). This issue was also raised in relation to interpreting statistics about the timing of amputation (immediate vs. delayed), and comparisons of treatments for fever in the Army and Navy (Edinburgh Medical and Surgical Journal 1813, p 458-459).
A writer in the Edinburgh Medical and Surgical Journal in 1813 stressed that, if one could assume the data to have been honestly assembled and presented by both sides, the only way out of the maze would be through "extensive comparative experiments" (Edinburgh Medical and Surgical Journal 1813).

During the 19th century there was gradual recognition that it is important to record the extent of uncertainty associated with estimates of treatment differences. In particular, Jules Gavarret, a mathematically inclined Parisian physician, pointed out the need to analyse treatment comparisons of sufficient size and to calculate the 'limits of oscillation' associated with statistical estimates of treatment differences (Gavarret 1840). However, this practice did not really become widely adopted until the second half of the 20th Century (see Explanatory Essay 3.2).

< Previous Essay | Next Essay >

References

Black W (1789). An arithmetical and medical analysis of the diseases and mortality of the human species. London: J Johnson.

Boylston AW (2010). Thomas Nettleton and the dawn of quantitative assessments of the effects of medical interventions. JLL Bulletin: Commentaries on the history of treatment evaluation (www.jameslindlibrary.org).

Boylston AW (2012). Defying providence: smallpox and the forgotten 18th century revolution.

Charleton R (1770). An inquiry into the efficacy of warm bathing in palsies. Oxford: Clarendon.

Clifton F (1732). The state of physick, ancient and modern, briefly considered: with a plan for the improvement of it. London printed by W Bowyer, for John Nourse without Temple-Bar, 1732.

Edinburgh Medical and Surgical Journal (1813);9:458-459.

Falconer W (1807). A practical dissertation on the medicinal effects of the Bath waters. London and Bath: Robinson & Crutwell.

Fordyce G (1793). An attempt to improve the evidence of medicine. Transactions of a society for the Improvement of medical and chirurgical knowledge. London: J Johnson.

Gavarret LDJ (1840). Principes généraux de statistique médicale: ou dèveloppement des règles qui doivent provider à son emploi. Paris: Bechet Jeune & Labé.

Haygarth J (1784). An inquiry on how to prevent the small-pox. Chester and London: Monk and Johnson.

Huth EJ (2005). Quantitative evidence for judgments on the efficacy of inoculation for the prevention of smallpox: England and New England in the 1700s. JLL Bulletin: Commentaries on the history of treatment evaluation (www.jameslindlibrary.org).

Jurin J (1724). A letter to the learned Dr. Caleb Cotesworth, F. R. S. of the College of Physicians, London, and physician to St. Thomas's Hospital; containing a comparison between the danger of the natural small pox, and that given by inoculation. Philosophical Transactions of the Royal Society of London (1722 – 1723), 32:213-227.

Lind (1772). A treatise of the scurvy. In three parts. Containing an inquiry into the nature, causes and cure, of that disease. Together with a critical and chronological view of what has been published on the subject. (Third Edition) London: Printed for S Crowder, D Wilson and G Nicholls, T Cadell, T Beket and Co.

Maehle A-H(2011). Four early clinical studies to assess the effects of Peruvian bark. JLL Bulletin: Commentaries on the history of treatment evaluation (www.jameslindlibrary.org).

McCausland R (1783) . Facts and observations on different medical subjects. Medical Commentaries viii (1781/82), 247-96.

McGrigor J (1801). Account of diseases of the 88th Regiment, during their passage to India, and at Bombay, from December 1798 till June 1800. Annals of Medicine, vol I Lustrum II, 1:353-369.

McGrigor J (1815). Sketch of the medical histories of the British Armies in the Peninsula of Spain and Portugal, during the late campaigns. Med.-chir. Trans. 6:381-489.

Millar J (1798). Observations on the conduct of the war. An appeal to the people of Great Britain. London: for the author.

Mills T (1813). An essay on the utility of bloodletting in fever. Dublin : J. Barlow for Gilbert and Hodges, London: Longman, Hurst, Rees, Orme and Brown.

Nettleton T (1722a). A letter from Dr. Nettleton, physician at Halifax in Yorkshire, to Dr. Whitaker concerning

the inoculation of the small pox.. Phil. Trans. 32, 1722 No. 370. p 35-48. Also published as an Account of the success of inoculating the small-pox in a letter to Dr. William Whitaker. Printed by S. Palmer for J. Batley London.

Nettleton T (1722b). A letter to James Jurin dated June 16, 1722. Published as part of a letter from Dr. Nettleton Physician, a letter from the same learned and ingenious gentleman concerning his further progress in inoculating the small pox. Philosophical Transactions, vol. 32, 1722 no. 370. p 49-52. Original in Royal Society Archives, Early Letters vol 1.

Nettleton T (1722c). Part of a letter from Dr. Nettleton, physician at Halifax, to Dr. Jurin, R.S. Secr. Concerning the inoculation of the smallpox, and the mortality of that distemper in the natural way. Philosophical Transactions of the Royal Society of London. 32:209-212.

Reide TD (1793). A view of the diseases of the Army in Great Britiain, America, the West Indies and on board of King's ships. London: Johnson.

Rollo J (1781). Observations on the diseases which appeared in the Army on St Lucia…, Barbados. Orderson for the author.

Rollo J (1801). A short account of the Royal Artillery Hospital at Woolwich. London: Mawman.

Tröhler U (2010). The introduction of numerical methods to assess the effects of medical interventions during the 18th century: a brief history. JLL Bulletin: Commentaries on the history of treatment evaluation (www.jameslind.library)

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

3.2 Quantifying uncertainty in treatment comparisons

Cite as: The James Lind Library 3.2 Quantifying uncertainty in treatment comparisons (http://jameslindlibrary.org/essays/3-2-quantifying-uncertainty-in-treatment-comparisons/)

> Chance may affect the results of a study if too few outcomes have been observed to yield reliable estimates of treatment effects. Small studies in which few outcome events occur are usually not informative and the results are sometimes seriously misleading.

To assess the role that chance may have played in the results of fair tests, researchers use 'tests of statistical significance'. When statisticians and others refer to 'significant differences' between treatments, they are usually referring to statistical significance. Statistically significant differences between treatments are not necessarily of any practical importance. But tests of statistical significance are important nevertheless because they help us to avoid mistaken conclusions that real differences in treatments exist when they don't – sometimes referred to as Type I errors. It is also important to take account of a sufficiently large number of outcomes of treatment to avoid a far more common danger – concluding that there are no differences between treatments when in fact there are. These mistakes are sometimes referred to as Type II errors.

Awareness of the importance of taking account of the play of chance began during the 19th century (list relevant records). Thomas Graham Balfour, for example, interpreted the results of his test of claims that belladonna could prevent the orphans under his care developing scarlet fever (Balfour 1854). Two out of 76 boys allocated to receive belladonna developed scarlet fever compared with 2 out of 75 boys who did not receive the drug. Balfour noted that "the numbers are too small to justify deductions as to the prophylactic power of belladonna". If more of the boys had developed scarlet fever, Balfour might have been able to reach a more confident conclusion about the possible effects of belladonna. Instead, he simply noted that 4 cases of scarlet fever among 151 boys was too small a number to reach a confident conclusion.

One approach that reduces the likelihood that we will be misled by chance effects involves estimating a range of treatment differences within which the real differences are likely to lie (Gavarret 1840; Huth 2006). These range estimates are known as confidence intervals. Repeating a treatment comparison is likely to yield varying estimates of the differential effects of treatments on outcomes, particularly if the estimates are based on small numbers of outcomes. Confidence intervals take account of this variation, and so they are more informative than mere tests of statistical significance, and thus more helpful in reducing the likelihood that we will be misled by the play of chance.

Statistical tests and confidence intervals – whether for analysis of individual studies, or in meta-analysis of a number of separate but similar studies – help us to take account of the play of chance and avoid concluding that treatment effects and differences exist when they don't, and don't exist when they do.

< Previous Essay | Next Essay >

References

Balfour TG (1854). Quoted in West C. Lectures on the Diseases of Infancy and Childhood. London, Longman, Brown, Green and Longmans, p 600.

Gavarret LDJ (1840). Principes gènèraux de statistique mèdicale: ou dèveloppement des règles qui doivent prèsider á son emploi. Paris: Bechet jeune & Labè.

Huth EJ (2006). Jules Gavarret's Principes Généraux de Statistique Médicale: a pioneering text on the statistical analysis of the results of treatments.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

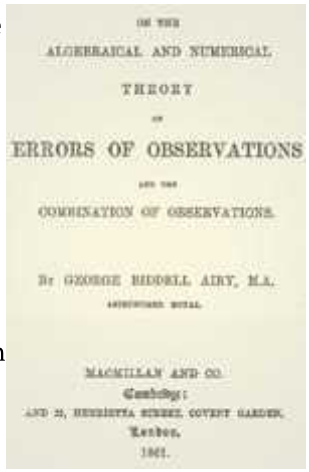
3.3 Reducing the play of chance using meta-analysis

Cite as: The James Lind Library 3.3 Reducing the play of chance using meta-analysis (http://jameslindlibrary.org/essays/3-3-reducing-the-play-of-chance-using-meta-analysis/)

Combining data from similar studies (meta-analysis) can help to provide more reliable estimates of treatment effects.

Systematic reviews of all the relevant, reliable evidence are needed for fair tests of medical treatments. To avoid misleading conclusions about the effects of treatments, people preparing systematic reviews must take steps to avoid biases of various kinds, for example, by taking account of all the relevant evidence and by avoiding biased selection from the available evidence.

Even though care may be taken to minimize biases in reviews, misleading conclusions about the effects of treatments may also result from the play of chance. Discussing separate but similar studies one at a time in systematic reviews may also leave a confused impression because of the play of chance. If it is both possible and appropriate, this problem can be reduced by combining the data from all the relevant studies, using a statistical procedure now known as 'meta-analysis'.


ON THE
ALGEBRAICAL AND NUMERICAL
THEORY
OF
ERRORS OF OBSERVATIONS
AND THE
COMBINATION OF OBSERVATIONS.
By GEORGE BIDDELL AIRY, M.A.
MACMILLAN AND CO.
Cambridge:
AND 23, HENRIETTA STREET, COVENT GARDEN,
London.
1861.

Most statistical techniques used today in meta-analysis derive from the work of the German mathematician Karl Gauss and the French mathematician Pierre-Simon Laplace during the first half of the 19th century. One of the fields in which their methods found practical application was astronomy: measuring the position of stars on a number of occasions often resulted in slightly different estimates, so techniques were needed to combine the estimates to produce an average derived from the pooled results. In 1861, the British Astronomer Royal, George Airy, published a 'textbook' for astronomers (Airy 1861) in which he described the methods used for this process of quantitative synthesis. Just over a century later, an American social scientist, Gene Glass, named the process 'meta-analysis' (Glass 1976).

An early medical example of meta-analysis was published in the British Medical Journal in 1904 by Karl Pearson (Pearson 1904; O'Rourke 2006), who had been asked by the government to review evidence on the effects of a vaccine against typhoid. Although methods for meta-analysis were developed by statisticians over the subsequent 70 years, it was not until the 1970s that they began to be applied more widely, initially by social scientists (Glass 1976), and then by medical researchers (Stjernswärd 1974; Stjernsward et al. 1976; Cochran et al. 1977; Chalmers et al. 1977; Chalmers 1979; Editorial 1980).

Meta-analysis can be illustrated using the logo of The Cochrane Collaboration. The logo illustrates a meta-analysis of data from seven fair tests. Each horizontal line represents the results of one test (the shorter the line, the more certain the result); and the diamond represents their combined results. The vertical line indicates the position around which the horizontal lines would cluster if the two treatments compared in the trials had similar effects; if a horizontal line crosses the vertical line, it means that particular test found no clear ('statistically significant') difference between the treatments. When individual horizontal lines cross the vertical 'no difference' line, it suggests that the treatment might either increase or decrease infant deaths. Taken together, however, the horizontal lines tend to fall on the beneficial (left) side

of the 'no difference' line. The diamond represents the combined results of these tests, generated using the statistical process of meta-analysis. The position of the diamond clearly to the left of the 'no difference' line indicates that the treatment is beneficial.

This diagram shows the results of a systematic review of fair tests of a short, inexpensive course of a steroid drug given to women expected to give birth prematurely. The first of these tests was reported in 1972. The diagram summarises the evidence that would have been revealed had the available tests been reviewed systematically a decade later, in 1981: it indicates strongly that steroids reduce the risk of babies dying from the complications of immaturity. By 1991, seven more trials had been reported, and the picture in the logo had become still stronger.

No systematic review of these trials was published until 1989 (Crowley 1989; Crowley et al 1990), so most obstetricians, midwives, and pregnant women did not realise that the treatment was so effective. After all, some of the tests had not shown a 'statistically significant' benefit, and maybe only these tests had been noticed. Because no systematic reviews had been done, tens of thousands of premature babies suffered, and many died unnecessarily. This is just one of many examples of the human costs that can result from failure to assess the effects of treatments in systematic, up-to-date reviews of fair tests, using meta-analysis to reduce the likelihood that the play of chance will be misleading. Furthermore, resources were wasted on unnecessary intensive care and research.

By the end of the 20th century it had become widely accepted that meta-analysis was an important element of fair tests of treatments, and that it helped to avoid incorrect conclusions that treatments had no effects when they were, in fact, either useful or harmful.

< Previous Essay | Next Essay >

References

Airy GB (1861). On the algebraical and numerical theory of errors of observations and the combination of observations. London: Macmillan.

Chalmers I (1979). Randomized controlled trials of fetal monitoring 1973-1977. In: Thalhammer O, Baumgarten K, Pollak A, eds. Perinatal Medicine. Stuttgart: Georg Thieme, 260-265.

Chalmers TC, Matta RJ, Smith H, Kunzler A-M. (1977). Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. New England Journal of Medicine 297:1091-1096.

Cochran WG, Diaconis P, Donner AP, Hoaglin DC, O'Connor NE, Peterson OL, Rosenoer VM (1977). Experiments in surgical treatments of duodenal ulcer. In: Bunker JP, Barnes BA, Mosteller F, eds. Costs, risks and benefits of surgery. Oxford: Oxford University Press, pp 176-197.

Crowley P (1989). Promoting pulmonary maturity. In: Chalmers I, Enkin M, Keirse MJNC, eds. Effective care in pregnancy and childbirth. Oxford: Oxford University Press, pp 746-762.

Crowley P, Chalmers I, Keirse MJNC (1990). The effects of corticosteroid administration before preterm delivery: an overview of the evidence from controlled trials. British Journal of Obstetrics and Gynaecology 97:11-25.

Editorial (1980). Aspirin after myocardialinfarction. Lancet 1:1172-3.

Glass GV (1976). Primary, secondary and meta-analysis of research. Educational Researcher 10, 3-8.

O'Rourke K (2006). An historical perspective on meta-anlysis: dealing quantatively with varying study results.

The James Lind Library.

Pearson K (1904). Report on certain enteric fever inoculation statistics. BMJ 3:1243-1246.

Stjernswärd J (1974). Decreased survival related to irradiation postoperatively in early operable breast cancer. Lancet 2:1285-1286.

Stjernswärd J, Muenz LR, von Essen CF (1976). Postoperative radiotherapy and breast cancer. Lancet 1:749.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

4.0 Bringing it all together for the benefit of patients and the public

Cite as: The James Lind Library 4.0 Bringing it all together for the benefit of patients and the public (http://jameslindlibrary.org/essays/4-0-bringing-it-all-together-for-the-benefit-of-patients/)

> Improving reports of research and preparing and updating systematic reviews of reliable studies are essential foundations of effective health care.

Fair treatment comparisons avoid biases and reduce as far as possible the likelihood that users of research will be misled by the play of chance. These problems and their potential solutions have been discussed in earlier Explanatory Essays. However, even if the problems have been reduced as far as possible, health professionals, patients, policy makers and the public more generally may often find it difficult to make direct use of reports of research.

Often, this is because the quality of reports of research – whether individual studies or systematic reviews of them – leaves a great deal to be desired. Too often reports fail to provide important details about the design, conduct and analysis of research studies; adequate descriptions of who participated in them; what was done to participants; and what effects treatments had on outcome measures of importance to patients and others (EE 4.1 Altman 1994).

Very occasionally, a single well conducted and well reported study provides very strong evidence of the beneficial effects of an easily given treatment. For example, tens of thousands of people participated in a remarkable study that showed that an aspirin tablet could substantially reduce the risk of death among people who are experiencing heart attacks (ISIS-2 1988). It is very rare, however, that a single study provides such strong evidence. And it's important when reading reports of individual studies to ask what other evidence – published and unpublished – is relevant. This is why systematic reviews of as high a proportion as possible of the relevant evidence are required to inform treatment and policy choices.

Systematic reviews are necessary, but they too, are insufficient for informing decisions about treatments for individual patients and policies. Other important factors – needs, resources and priorities – need to be taken into account. And this is the point at which the art as well as the science of health care needs to be deployed for the benefit of patients and the public (EE 4.3 Chalmers 1993; Rothwell 2007).

< Previous Essay | Next Essay >

References

Altman D (1994). The scandal of poor medical research. BMJ 308:283-284.

Chalmers I (1993). The Cochrane Collaboration: preparing, maintaining and disseminating systematic reviews of the effects of health care. In: Warren KS, Mosteller F, eds. Doing more good than harm: the evaluation of health care interventions. Annals of the New York Academy of Sciences 703:156-163.

ISIS-2 (second International Study of Infarct Survival) Collaborative Group (1988). Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17 187 cases of suspected acute myocardial infarction: ISIS-2. Lancet 332:349–360

Rothwell P (2007). Treating individuals: from randomized controlled trials to personalized medicine. London: Lancet.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

4.1 Improving reports of research

Cite as: The James Lind Library 4.1 Improving reports of research (http://jameslindlibrary.org/essays/4-1-improving-reports-of-research/)

> High quality, complete reports of research are needed to provide maximum return on the public's substantial investment in research on the effects of treatments.

The Medical Research Council's randomised trial comparing bed rest alone with bed rest and streptomycin for treating pulmonary tuberculosis (MRC 1948) is renowned for several reasons. As far as the research methods used are concerned, it is because it introduced secure methods for assuring that the comparison groups would be similar (Chalmers 2010). However, another feature of the report of the study was that it was exceptionally clearly drafted. This reflected the care taken by the three members of the research team. One of them, Marc Daniels, went on to publish papers commenting on the inadequacy of many reports of research, and recommending reporting standards (Daniels 1950; 1951). Some years later, Austin Bradford Hill, one of Daniels' two senior colleagues, also offered guidance (Hill 1965).

It was not until the 1980s that formal surveys of the quality of reports of research began to reveal just how common were deficiencies (Hemminki 1981; 1982), and that remedies began to be suggested in proposed reporting standards (Chalmers TC et al. 1981; Ad Hoc Working Group 1987). The 1990s witnessed concerted international initiatives to improve the quality of reports of research (Standards of Reporting Trials Group 1994; The Consort Group 1996). In a BMJ editorial in 1994, Douglas Altman commented on "the scandal of poor medical research", we need less research, better research and research done for the right reasons (Altman 1994). Since then, he and hios colleagues in the Equator Network (www.equator-network.org) have created a library of health research reporting guidelines. Promoting adherence to these guidelines by researchers and journals remains a challenge.

< Previous Essay | Next Essay >

References

Ad Hoc Working Group for Critical Appraisal of the Medical Literature (1987). A proposal for more informative abstracts of clinical articles. Annals of Internal Medicine 106:598-604.

Altman (1994). The scandal of poor medical research. BMJ 308:283-284.

Chalmers I (2010). Why the 1948 MRC trial of streptomycin used treatment allocation based on random numbers. JLL Bulletin: Commentaries on the history of treatment evaluation (www.jameslindlibrary.org).

Chalmers TC, Smith H, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A (1981). A method for assessing the quality of a randomized control trial. Controlled Clinical Trials 2: 31-49

Daniels M (1950). Scientific appraisement of new drugs in tuberculosis. American Review of Tuberculosis 61:751-756.

Daniels M (1951). Clinical evaluation of chemotherapy in tuberculosis. British Medical Bulletin 7:320-326.

Hemminki E (1981). Quality of reports of clinical trials submitted by the drug industry to the Finnish and Swedish control authorities. European Journal of Clinical Pharmacology 19:157-165.

Hemminki E (1982). Quality of clinical trials – a concern of three decades. Methods of Information in Medicine 21:81-85.

Hill AB (1965). The reasons for writing. BMJ 2:870.

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

Standards of Reporting Trials Group (1994). A proposal for structured reporting of randomized controlled trials. JAMA 272:1926-31.

The CONSORT Group (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT Statement. JAMA 276:637-639.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

4.2 Preparing and maintaining systematic reviews of all the relevant evidence

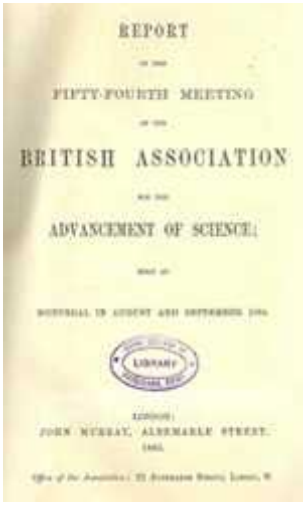
Cite as: The James Lind Library 4.2 Preparing and maintaining systematic reviews of all the relevant evidence (http://jameslindlibrary.org/essays/4-2-preparing-and-maintaining-systematic-reviews-of-all-the-relevant-evidence/)

> Unbiased, up-to-date systematic reviews of all the relevant, reliable evidence are needed to provide trustworthy evidence to inform practice and policy.

One of the twentieth century pioneers of fair tests of treatments, Austin Bradford Hill , noted that readers of reports of research want answers to four questions: 'Why did you start?', 'What did you do?', 'What did you find?', and 'What does it mean anyway?' (Hill 1965). The quality of the answer to Hill's last question is particularly important because this is the element of a research report which is most likely to influence actual choices and decisions about treatments.

Only very rarely will a single fair test of a treatment yield sufficiently strong evidence to provide a confident answer to the question 'What does it mean?' A fair test of a treatment is usually one of a number of tests addressing the same question. For a reliable answer to the question 'What does it mean?', then, it is important to interpret the evidence from a particular fair test in the context of a careful assessment of all the evidence from fair tests that have addressed the question concerned.

The president of the British Association for the Advancement of Science expressed the need to observe this principle more than a century ago:


REPORT

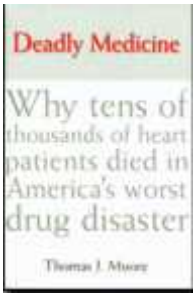
FIFTY-FOURTH MEETING

BRITISH ASSOCIATION

ADVANCEMENT OF SCIENCE;

> "If, as is sometimes supposed, science consisted in nothing but the laborious accumulation of facts, it would soon come to a standstill, crushed, as it were, under its own weight…. Two processes are thus at work side by side, the reception of new material and the digestion and assimilation of the old…The work which deserves, but I am afraid does not always receive, the most credit is that in which discovery and explanation go hand in hand, in which not only are new facts presented, but their relation to old ones is pointed out." (Rayleigh 1885)

Very few reports of fair tests of treatments discuss their results in the context of a systematic assessment of all the other relevant evidence (Clarke and Hopewell 2013). As a result, it is usually difficult for readers to obtain a reliable answer to the question 'What does it mean?' from reports of new research.



As noted in an earlier explanatory essay, embarking on new tests of medical treatments without first reviewing systematically what can be learnt from existing research is dangerous, wasteful and unethical (see Why comparisons must address genuine uncertainties). Reporting the results of new tests without interpreting new evidence in the light of systematic assessments of other relevant evidence is also dangerous because it results in delays in the identification of both useful and harmful treatments (Antman et al. 1992). For example, between the 1960s and the early 1990s, over 50 fair tests of drugs to reduce heart rhythm abnormalities in people having heart attacks were done before it was

realised that these drugs were killing people. Had each report assessed the results of new tests in the context of all the relevant evidence, the lethal effects of the drugs could have been identified a decade earlier, and many unnecessarily premature deaths could have been avoided (Clarke et al. 2014).

In an age of electronic publishing it should be possible to deal with the limitations found in most reports of new research (Chalmers and Altman 1999; Smith and Chalmers 2001). However, rather than basing conclusions about the treatments on one or a few individual studies, users of research evidence are increasingly turning for reliable information to up-to-date, systematic reviews of all relevant, reliable evidence, because these are increasingly recognised as providing the best basis for conclusions about the effects of medical treatments.

Just as it is important to take steps to avoid being misled by biases and the play of chance in planning, conducting, analysing and interpreting individual fair tests of treatments, so also must similar steps be taken in planning, conducting, analysing and interpreting systematic reviews. This entails:

- specifying the question to be addressed by the systematic review
- defining eligibility criteria for studies to be included
- identifying (all) potentially eligible studies
- applying eligibility criteria in ways that limit bias
- assembling as high a proportion as possible of the relevant information from the studies
- analysing this information, if appropriate and possible, using meta-analysis and a variety of analyses
- preparing a structured report

One manifestation of the increasing recognition of the crucial importance of systematic reviews for assessing the effects of treatments has been the rapid evolution of methods to improve the reliability of reviews themselves. The first edition of a book entitled Systematic Reviews was less than 100 pages long (Chalmers and Altman 1995): only six years later, the second edition weighed in at nearly 500 pages and included rapidly evolving strategies for increasing the information obtained from research (Egger et al. 2001).

There continue to be important developments in the methods used for preparing systematic reviews, including those needed to identify unanticipated effects of treatments and for incorporating the results of research describing and analysing the experiences of people giving and receiving treatments (Thomas et al 2004; Jefferson et al 2014).

< Previous Essay | Next Essay >

References

Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC (1992). A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. JAMA 268:240-48.

Chalmers I, Altman DG (1995). Systematic Reviews. London: BMJ Publications.

Chalmers I, Altman DG (1999). How can medical journals help prevent poor medical research? Some opportunities presented by electronic publishing. Lancet 353:490-493.

Clarke M, Hopewell S (2013). Many reports of randomised trials still don't begin or end with a systematic review of the relevant evidence. J Bahrain Med Soc 24: 145-148.

Clarke M, Brice A, Chalmers I (2014). Accumulating research: a systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. PLoS ONE 9(7): e102670. doi:10.1371/journal.pone.0102670.

Egger M, Davey Smith G, Altman D (2001). Systematic Reviews in Health Care: meta-analysis in context. 2nd Edition of Systematic Reviews. London: BMJ Books.

Hill AB (1965). Cited in 'The reasons for writing'. BMJ 4:870.

Jefferson T, Jones MA, Doshi P, Del Mar CB, Hama R, Thompson MJ, Spencer EA, Onakpoya I, Mahtani KR, Nunan D, Howick J, Heneghan CJ (2014). Neuraminidase inhibitors for preventing and treating in?uenza in healthy adults and children. Cochrane Database of Systematic Reviews 2014, Issue 4. Art. No.: CD008965. DOI:10.1002/14651858.CD008965.pub4

Rayleigh, Lord (1885). Address by the Rt. Hon. Lord Rayleigh. In: Report of the fifty-fourth meeting of the British Association for the Advancement of Science; held at Montreal in August and September 1884, London: John Murray.

Smith R, Chalmers I (2001). Britain's gift: a 'Medline' of synthesized evidence. BMJ 323:1437-1438.

Thomas J, Harden A, Oakley A, Oliver S, Sutcliffe K, Rees R, Brunton G, Kavanagh J (2004). Integrating qualitative research with trials in systematic reviews BMJ 328:1010-1012

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT

James Lind Library

Illustrating the development of fair tests of treatments in health care

4.3 Using the results of up-to-date systematic reviews of research

Cite as: The James Lind Library 4.3 Using the results of up-to-date systematic reviews of research (http://jameslindlibrary.org/essays/4-3-using-the-results-of-up-to-date-systematic-reviews-of-research/)

> All research has been done in the past, but the results of research need to be used today and tomorrow to inform decisions in health care. Trustworthy evidence from research is necessary, but not sufficient, to improve the quality of health care.

Over recent years it has been realised increasingly that systematic reviews of research are needed to provide fair tests of treatments. This trend has been reflected in a rapid increase in the numbers of reports of systematic reviews being published on paper and electronically (Bastian et al. 2010). Sometimes reviews will show that no reliable evidence exists, and this is one of their most important functions. Similarly, they may sometimes confirm that reliable evidence is limited to a single study, and here, too, it is important to make this situation explicit.

Systematic reviews of research are being used widely (i) to inform clinical practice, often through clinical practice guidelines; (ii) to assess which medical treatments are cost-effective; (iii) to shape the agenda for additional research; and (iv) to meet the needs of patients for reliable information about the effects of treatments.

Although these developments show that the importance of systematic reviews has been accepted by those who are trying to improve access to the evidence needed to inform choices in health care (Smith and Chalmers 2001), there is still a long way to go. Many thousands of systematic reviews will be needed to cover existing research evidence, and then kept up to date as new evidence emerges. Indeed, one journal editor has suggested that there should be a moratorium on all new research until we've caught up with what existing evidence can tell us (Bausell 1993).

Those responsible for disbursing funds for research must ensure that resources are provided to cope with this backlog, and that new studies are only supported if systematic reviews of existing evidence have shown that additional studies are necessary, and that they have been designed to take account of the lessons from previous research. If journal editors are to serve their readers better, they must follow the lead of The Lancet and ensure that reports of new studies make clear what contribution new evidence has made to an up-to-date systematic review of all the relevant evidence (Young and Horton 2005).

The increased availability of up-to-date, systematic reviews is improving the quality of information about the effects of treatments, but the conclusions of systematic reviews should not be accepted uncritically. Different reviews purportedly addressing the same question about treatments sometimes arrive at different conclusions. Their authors are human and we need to be aware that they may select, analyse and present evidence in ways that support their prejudices and interests. The continuing evolution of reliable methods for preparing and maintaining systematic reviews will help to address this problem, but they cannot be expected to abolish it.

Systematic reviews are necessary but insufficient for informing decisions about treatments for individual patients and policies. Other important factors – needs, resources and priorities – need to be taken into account (Chalmers 1993; Rothwell 2007). And this is the point at which the art as

well as the science of health care needs to be deployed for the benefit of patients and the public (Evans et al. 2011; www.testingtreatments.org).

< Previous Essay

References

Bastian H, Glasziou P, Chalmers I (2010). Seventy-five trials and eleven systematic reviews a day: How will we ever keep up? PLoS Medicine 7:e1000326.

Bausell BB (1993). After the meta-analytic revolution. Evaluation and the Health Professions 16:3-12

Chalmers I (1993). The Cochrane Collaboration: preparing, maintaining and disseminating systematic reviews of the effects of health care. In: Warren KS, Mosteller F, eds. Doing more good than harm: the evaluation of health care interventions. Annals of the New York Academy of Sciences 1993;703:156-163.

Evans I, Thornton H, Chalmers I, Glasziou P (2011). Testing Treatments. London: Pinter and Martin.

Rothwell P (2007). Treating individuals: from randomized controlled trials to personalized medicine. London: Lancet.

Smith R, Chalmers I (2001). Britain's gift: a 'Medline' of synthesized evidence. BMJ 323:1437-1438.

Young C, Horton R (2005). Putting clinical trials into context. Lancet 366:107-8.

BROWSE

Topics
Essays
Records
Articles

COPYRIGHT