



Why comparisons are essential

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. 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 (Cumings 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

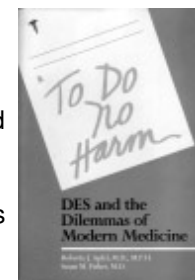
Treatment comparisons based on impressions, or relatively restricted analyses, only provide reliable information in the

rare circumstances when treatment effects are dramatic ([click here to list relevant records](#), [Glasziou et al. 2007](#)). Examples include opium for pain relief ([Tibi 2005](#)), hygiene for preventing tetanus (lockjaw) in newborn babies (Schleisner 1849), chloroform for anaesthesia, 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](#)), streptomycin for tuberculous meningitis ([MRC 1948](#)), adrenaline for life-threatening allergic reactions (McLean-Tooke et al. 2003), and genetically-designed drugs for some rare forms of leukaemia (Druker et al. 2001). Most medical treatments don't have such dramatic effects as these, 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

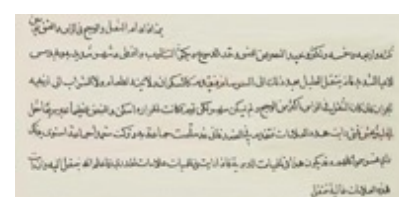
Sometimes giving different treatments at more or less the same time may involve giving a patient different treatments one after the other – a so-called crossover test ([Martini 1932](#); [click here to list relevant records](#)). An early example of a crossover test was reported in 1786 by Dr Caleb Parry 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 9th century](#)).



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](#).

Cite As:

Editorial commentary (2007). Why comparisons are essential. The James Lind Library (www.jameslindlibrary.org).