

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

Avoiding biased comparisons:

Differences between the people compared

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.

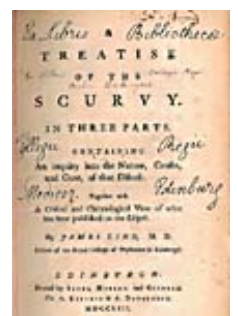
The 18th century surgeon William Cheselden was aware of the 'dissimilar groups' problem when surgeons were comparing their respective mortality rates after operations to remove bladder stones. Cheselden pointed out that it was important to take account of the ages of the people treated by different surgeons. 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 setup 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 had responsibility, 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 [Hamilton 1816](#); [MRC 1944](#)), or by

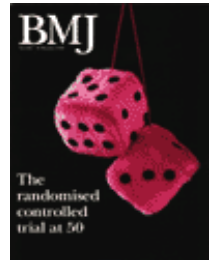


drawing lots ([Colebrook 1929](#)) – for example, using dice ([Doull et al. 1931](#)), coloured beads ([Theobald 1937](#)), or random sampling numbers ([Bell 1941](#); [MRC 1948](#); [MRC 1950](#); [MRC 1951](#)). This 'random allocation' is the sole, but crucially important, feature of the category 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 can not 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.



As illustrated in the essay available by clicking [here](#), casting or drawing lots is a time-honoured way of making fair decisions. 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 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](#)).



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 these 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). 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 randomized 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.

Cite as: Editorial commentary (2007). Differences between the people compared. The James Lind Library (www.jameslindlibrary.org).

Show JLL records: illustrating [control of allocation bias](#) and [crossover test](#)

Next essay: [Differences in the way treatment outcomes are assessed](#)

Select other 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.
- Hamilton AL (1816). Dissertatio Medica Inauguralis De Synocho Castrensi (Inaugural medical dissertation on camp fever). Edinburgh: J Ballantyne.
- 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. Casting and drawing lots. The James Lind Library (www.jameslindlibrary.org).
- Theobald GW (1937). Effect of calcium and vitamin A and D on incidence of pregnancy toxæmia. Lancet 2: 1397-1399.

[Home](#)

[Contents](#)

[Comments welcome](#)