ENDGAMES

STATISTICAL QUESTION

Equivalence trials

Philip Sedgwick *reader in medical statistics and medical education*

Centre for Medical and Healthcare Education, St George's, University of London, Tooting, London, UK

Researchers evaluated the efficacy of 4% dimeticone lotion for the treatment of head louse infestation. A randomised controlled equivalence trial was performed. Control treatment was 0.5% phenothrin liquid, the standard treatment in the United Kingdom at the time of the trial. Treatments were applied twice, seven days apart, with dimeticone lotion for eight hours or overnight, and phenothrin liquid for 12 hours or overnight.[1]

The primary outcome was proportion of participants cured of infestation after the second application, regardless of whether reinfestation occurred later. The trial was designed to demonstrate therapeutic equivalence with an equivalence margin of 20%. Participants were young people (4-18 years) and adults with active head louse infestation. In total, 127 participants were allocated to the intervention (dimeticone lotion) and 125 to control (phenothrin liquid).

Analysis by intention to treat indicated that 89 of 127 (70%) participants treated with dimeticone were cured compared with 94 of 125 (75%) treated with phenothrin (difference −5%, 95% confidence interval −16% to 6%). Analysis by per protocol indicated that 84 of 121 (69%) participants were cured with dimeticone compared with 90 of 116 (78%) with phenothrin, a difference of −8% (−19% to 3%).

Which of the following statements, if any, are true?

a) The control treatment of 0.5% phenothrin liquid is described as an active control

b) The null hypothesis for the equivalence trial stated no difference between treatments in cure rate in the population from where the participants were selected

c) Therapeutic equivalence in cure rate to within 20% was demonstrated between treatments

Answers

Answers *a* and *c* are true, whereas *b* is false.

The aim of the trial was to investigate whether 4% dimeticone lotion and 0.5% phenothrin liquid, each applied twice seven days apart, were therapeutically equivalent. The primary outcome was percentage cured after the second application, regardless of whether reinfestation occurred later. Equivalence trials compare a new treatment for a disease or condition with

an existing treatment—usually the standard one. Treatments that have already been shown to be effective are referred to as active controls when used as a control treatment in a trial (*a* is true).

Classic randomised controlled trials aim to establish whether a new treatment or therapeutic regimen is better than an existing one or placebo. Sometimes referred to as superiority trials, described in a previous question,[2] they are analysed using the traditional approach of statistical hypothesis testing.[3] The null hypothesis starts at the position of equipoise—that is, the treatments are therapeutically equivalent in the population from which the trial participants were selected. The purpose is to establish whether the data provide sufficient evidence to reject the null hypothesis in favour of the alternative, which states that the treatments are not therapeutically equivalent in the population—the new treatment is either superior or inferior to the comparator (standard treatment or placebo). Hypothesis testing involves a statistical significance test, with a resulting P value that determines whether there is sufficient evidence to reject the null hypothesis in favour of the alternative. A 95% confidence interval is usually also derived to provide an interval estimate of the potential difference between treatments in the population.

If the null hypothesis in a superiority trial is not rejected, it cannot be concluded that the treatments are therapeutically equivalent, only that there was insufficient evidence to reject the null hypothesis in favour of the alternative. Although a trial may fail to find a significant difference between treatments, it does not mean that one does not exist. The trial participants were a single sample and may not have been representative of the population owing to sampling error. Another sample may have given different results. An equivalence trial was needed to demonstrate therapeutic equivalence between dimeticone and phenothrin.

In an equivalence trial the statistical null and alternative hypotheses are a reversal of those for a superiority trial. In the example above, the null hypothesis stated that in the population from which the participants were selected, dimeticone was not therapeutically equivalent to phenothrin—that is, dimeticone was inferior or superior to phenothrin (*b* is false). The aim of

p.sedgwick@sgul.ac.uk

the trial was to establish whether the data provided sufficient evidence to reject the null hypothesis in favour of the alternative, which stated that in the population dimeticone and phenothrin were therapeutically equivalent. However, unlike superiority trials, statistical significance testing and P values are not used to analyse equivalence trials. Analysis is based on confidence intervals.

It would not have been possible to prove that dimeticone and phenothrin were exactly equivalent therapeutically. Therefore, a margin of equivalence in the primary outcome based on clinical reasons was proposed. An equivalence margin of 20% was suggested—that is, dimeticone and phenothrin would be considered equivalent if the difference between them in cure rate was no larger than 20%. Therefore, the proposed range of equivalence between treatments was −20% to 20%, depending on whether dimeticone had a smaller or larger cure rate than phenothrin.

The proposed equivalence range is for the true difference between treatments in the population, as estimated by the observed data in the trial. The confidence interval, typically a 95% one, is used to provide an interval estimate of the potential difference between treatments in the population. Therefore, if the confidence interval based on the observed data lies within the equivalence range, treatments are considered therapeutically equivalent. Equivalence trials will often be analysed using an intention to treat and per protocol approach, both of which have been described in previous questions.[4][5] The 95% confidence interval for the difference between treatments in cure rate (dimeticone minus phenothrin) was −16% to 6% using an intention to treat analysis, and −19% to 3% using a per protocol approach. Both confidence intervals lie entirely within the equivalence range. Therefore, on the basis of either analysis the treatments were therapeutically equivalent in cure rate to within 20% (*c* is true). Any difference between treatments in cure rate was considered to be of no clinical relevance. If the 95% confidence interval for the difference between treatments in cure rate did not lie entirely within the equivalence range—for example, if one limit of the confidence interval extended below −20% or above 20%, then equivalence could not have been assumed.

As described above, if a superiority trial fails to demonstrate superiority of one treatment over another, it is not possible to infer therapeutic equivalence between treatments under the null hypothesis. It would also be inappropriate to analyse a superiority trial post hoc as an equivalence trial in an attempt to demonstrate therapeutic equivalence unless both superiority and equivalence analyses were explicitly stated in the trial protocol.

Equivalence trials are useful because it is not always possible to develop new drugs that are sufficiently more effective than the standard treatment. However, it may be beneficial to patients to develop new treatments that are therapeutically equivalent, not least because they would provide alternative or second line treatments. New treatments might also be cheaper, have fewer side effects, and be easier to administer.

Non-inferiority trials, described in a previous question,[6] tend to be used instead of equivalence trials to demonstrate therapeutic similarity between treatments. Equivalence and non-inferiority trials are similar. The alternative hypothesis for a non-inferiority trial states that the efficacy of a new treatment is similar or superior, but no worse (by more than a pre-stated equivalence margin), than that of the standard treatment. However, the alternative hypothesis for an equivalence trial does not consider superiority of the new treatment over the standard one, only that the new treatment is not different from the standard one in efficacy by more than the pre-stated equivalence margin. The concept of equivalence is more often used in so called bioequivalence trials, the aim of which is to establish whether two or more formulations of a drug containing the same active ingredient give comparable blood concentrations with a pre-stated equivalence margin.

Competing interests: None declared.

1 Burgess IF, Brown CM, Lee PN. Treatment of head louse infestation with 4% dimeticone lotion: randomised controlled equivalence trial. *BMJ* 2005;330:1423-5.
2 Sedgwick P. Superiority trials. i 2011;342:d2981.
3 Sedgwick P. Statistical hypothesis testing. *BMJ* 2010;340:c2059.
4 Sedgwick P. Analysis by intention to treat. *BMJ* 2011;342:d2212.
5 Sedgwick P. Analysis by per protocol. *BMJ* 2011;342:d2330.
6 Sedgwick P. Non-inferiority trials. *BMJ* 2011;342:d3253.

Cite this as: *BMJ* 2013;346:f184

© BMJ Publishing Group Ltd 2013