ENDGAMES

STATISTICAL QUESTION

Clinical significance versus statistical significance

Philip Sedgwick *reader in medical statistics and medical education*

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

The effectiveness of topical chloramphenicol in preventing wound infection after minor dermatological surgery was evaluated. A randomised placebo controlled double blind superiority trial was performed. The intervention was a single application of topical chloramphenicol ointment applied to the sutured wound immediately after suturing. Participants were patients with high risk sutured wounds who had undergone minor surgery.[1]

The primary outcome was infection on the agreed day of removal of sutures or sooner if the patient re-presented with a perceived infection. The required sample size was based on a projected infection rate of 10% in the placebo group. The smallest effect of clinical interest was an absolute decrease in the incidence of infection of 5%. To achieve this difference with a power in excess of 80% and a critical level of significance of 0.05, 473 patients were needed in each treatment group. In total, 972 patients were recruited and randomised to topical chloramphenicol ointment (n=488) or placebo (n=484).

The proportion of participants with an infection in the topical chloramphenicol group was statistically significantly lower than for placebo (6.6% *v* 11.0%; difference −4.4%, 95% confidence interval −7.9% to −0.8%; P=0.010).

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

a) The derived sample size was based on clinical significance

b) The derived sample size was based on statistical significance

c) It can be inferred that the reduction in infection rate for the intervention compared with placebo was clinically significant because it was statistically significant

Answers

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

The aim of the trial was to evaluate the effectiveness of topical chloramphenicol compared with placebo in preventing wound infection after minor dermatological surgery. Before starting the trial it was necessary to calculate the optimal sample size. The importance of having an optimal sample size in a clinical trial has been described in a previous question.[2] The required number of participants was based on clinical and statistical

significance (*a* and *b* are true). The infection rate for the placebo group was predicted to be 10%. The smallest effect of clinical interest was considered to be an absolute decrease in incidence of infection of 5%. If the intervention group reduced the infection rate by 5% or more compared with placebo, then topical chloramphenicol would be considered clinically effective. To see this difference in effectiveness and demonstrate it as statistically significant, with power in excess of 80% and at a critical level of significance of 0.05, 473 patients would need to be recruited to each treatment group.

The proportion of participants with an infection in the topical chloramphenicol group was lower than for placebo (6.6% *v* 11.0%). The reduction in risk was statistically significant (difference −4.4%, 95% confidence interval −7.9% to −0.8%; P=0.010). The inference of statistical significance was based on the P value, which was derived from a statistical hypothesis test. The P value measured the strength of the evidence in support of the null hypothesis. The trial was designed as a superiority one. Superiority trials have been described in a previous question.[3] The null hypothesis stated that there was no difference in infection rate between the intervention and placebo. The critical level of significance for statistical testing was set at 0.05 (5%). Because P=0.010, there was little evidence to support the null hypothesis and it was rejected in favour of the alternative. The inference was that the incidence of infection was statistically significantly reduced in the intervention group.

Statistical significance implies that the difference seen in the sample also exists in the population. Clinical significance implies that the difference between treatments in effectiveness is clinically important, and it is possible that clinical practice will change if such a difference is seen. Statistical significance is used to inform clinical significance. However, clinical significance and statistical significance are often confused. The terms are often used interchangeably, although one does not necessarily imply the other. Researchers sometimes infer that the effectiveness of a treatment is clinically significant because the difference between treatments is statistically significant. However, clinical significance cannot necessarily be inferred from statistical significance (*c* is false), and statistical significance cannot be inferred from clinical significance.

p.sedgwick@sgul.ac.uk

The difference in incidence of infection was a reduction of 4.4% for the intervention compared with placebo, which was statistically significant (P=0.010). However, although the reduction in infection was statistically significant, the researchers concluded that it was not clinically significant. Their conclusion was justified because the reduction in incidence of infection was less than the smallest effect of clinical interest (5%) (*c* is false).

The smallest effect of clinical interest was a 5% absolute reduction in incidence of infection. Obviously, an absolute reduction in incidence larger than 5% would have been regarded as clinically significant. However, the smallest effect of clinical interest may not exist for the population of patients after minor dermatological surgery. That is, the difference in rate of infection that would be seen between treatments groups if applied to the entire population may be less than 5%. However, if the smallest effect of clinical interest does exist for the population, the probability that this effect will be seen in the trial needs to be maximised. To do so, an optimal sample size was needed. This underlies the concept of statistical power, as described in a previous question.[4] It is obviously desirable for power to be as large as possible. However, the implication of increasing statistical power is that it results in an increased sample size. Therefore, a compromise between power and sample size is usually achieved. The power was set to at least 80% in the above trial, this being the minimum generally recommended when calculating sample size in clinical trials.

In the above trial, 435 participants were needed in each treatment group to achieve 80% power. The researchers recruited and randomised more participants than this to each treatment group. A disadvantage of increasing sample size, and therefore the power of detecting the smallest effect of clinical interest, is that differences between treatment groups less than the smallest effect of clinical interest are more likely to be found to be statistically significant. In effect, the trial was overpowered. This is the most likely explanation for why the above trial found a statistically significant difference that was not clinically significant. Overpowered trials have been described in a previous question.[4]

In addition to the terms clinical significance and statistical significance, a further type of significance is proposed here—"patient significance." It is becoming increasingly popular to have patient groups involved in research and the development of clinical trials. Presumably in the trial above the application of topical chloramphenicol ointment to the sutured wound immediately after suturing was of little inconvenience to patients, especially if it reduced the risk of infection. The smallest effect of clinical interest was based on clinical expertise and experience. However, in other trials, any proposed changes in outcome based on clinical expertise and experience may not be significant to patients, especially if the intervention requires a considerable amount of time and effort on behalf of the patient.

Competing interests: None declared.

1   Heal CF, Buettner PG, Cruickshank R, Graham D, Browning S, Pendergast J, et al. Does single application of topical chloramphenicol to high risk sutured wounds reduce incidence of wound infection after minor surgery? Prospective randomised placebo controlled double blind trial. *BMJ* 2009;338:a2812.
2   Sedgwick P. Sample size: how many participants are needed in a trial? *BMJ* 2013;346:f1041.
3   Sedgwick P. Superiority trials. *BMJ* 2011;342:d2981.
4   Sedgwick P. The importance of statistical power. *BMJ* 2013;347:f6441.

Cite this as: *BMJ* 2014;348:g2130