BMJ

BMJ 2013;346:f3212 doi: 10.1136/bmj.f3212 (Published 17 May 2013)

ENDGAMES

STATISTICAL QUESTION

P values or confidence intervals?

Philip Sedgwick *reader in medical statistics and medical education*

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

The effectiveness of topical chloramphenicol in preventing wound infection after minor dermatological surgery was evaluated. Researchers performed a randomised placebo controlled double blind superiority trial. The intervention was a single application of topical chloramphenicol ointment to the sutured wound immediately after suturing. The placebo treatment was a single application of paraffin ointment. In total, 972 minor surgery patients with high risk sutured wounds were recruited. Trial participants were randomised to topical chloramphenicol ointment (n=488) or placebo (n=484). The primary outcome was infection on the agreed day of removal of sutures or sooner if the patient re-presented with a perceived infection.[1]

The critical level of significance for statistical testing was set at 0.05 (5%). The proportion of participants with an infection was significantly lower in the chloramphenicol group than in the placebo group (6.6% *v* 11.0%; difference −4.4%, 95% confidence interval −7.9% to −0.8%; P=0.010). The researchers concluded that the application of a single dose of topical chloramphenicol to high risk sutured wounds after minor surgery produced a statistically significant yet moderate absolute reduction in infection rate.

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

a) The P value provides a direct statement about the size of the difference between groups in the proportion of patients with wound infection

b) The P value provides a direct statement about the direction of the difference between groups in the proportion of patients with wound infection

c) The P value provides a dichotomous test of significance of the statistical hypotheses

d) The 95% confidence interval provides a dichotomous test of significance of the statistical hypotheses

Answers

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

The derivation of the P value (P=0.010) was based on statistical hypothesis testing, described in a previous question.[2] Because the trial had a superiority design the alternative hypothesis was two sided.[3] The statistical null hypothesis stated that in the population from which the sample was taken there was no difference between intervention and placebo in the proportion of minor surgery patients with wound infection. The alternative hypothesis stated that, in the population, intervention when compared with placebo would result in either a larger or smaller proportion of patients with wound infection.

In the above example, the P value was less than the critical level of significance of 0.05 (5%); this critical level of significance is chosen by convention. Therefore, the null hypothesis was rejected in favour of the alternative one and the difference between treatment groups in percentage wound infection deemed significant. However, although the difference between treatment groups was significant, the P value alone cannot inform any direct statement about the size of the difference between groups in proportion of minor surgery patients with wound infection (*a* is false). Nonetheless, smaller P values would indicate larger differences between treatment groups. Furthermore, the P value provides no indication of the direction of the effect studied—whether the infection rate was higher or lower for the intervention group compared with placebo (*b* is false). As such, statistical hypothesis testing based on a critical level of significance is a dichotomous test (*c* is true).

The 95% confidence interval, described in a previous question,[4] is an interval estimate for the population parameter; it represents the uncertainty of the sample in estimating the population parameter as a result of sampling error.[5] In the example above, the population parameter is the mean difference between intervention and placebo in the proportion of minor surgery patients with wound infection if the treatments were applied to the entire population. The 95% confidence interval (−7.9% to −0.8%) includes the population parameter with a probability of 0.95 (95%). The confidence interval is based on the hypothetical situation of repeating the above study an infinite number of times and under the same conditions. Samples would be randomly selected from the population and therefore include different patients. In turn, the mean difference between treatments in the proportion of patients with wound infection would differ between samples. For each sample, a 95% confidence interval for the mean difference between treatments

p.sedgwick@sgul.ac.uk

For personal use only: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ: first published as 10.1136/bmj.f3212 on 17 May 2013. Downloaded from http://www.bmj.com/ on 16 October 2023 at CzechELib project (Charles University). Protected by copyright.

could be calculated. The 95% confidence interval would differ between samples, and in 95 of 100 samples the calculated 95% confidence interval would include the population parameter.

Conclusions about significance can be made on the basis of the 95% confidence interval.[6] Because the 95% confidence interval for the mean difference between treatment groups in the proportion of patients with wound infection excludes zero, then the test of the statistical null and alternative hypotheses described above is significant at the 5% level—P is less than 0.05 (5%). The null hypothesis is rejected in favour of the alternative, with the conclusion that there is a significant difference between treatment groups. Therefore, similar to the P value, the 95% confidence interval can be used to perform a dichotomous test of statistical hypothesis testing (*d* is true). However, in contrast to the P value, the 95% confidence interval indicates the direction and size of the effect studied.

The limits of the confidence interval are below zero, so it is implied that a smaller proportion of minor surgery patients had a wound infection in the intervention group than in the placebo group. As an interval estimate, the confidence estimate provides a range of values within which the population parameter is contained, with a probability of 0.95. Even if the 95% confidence interval for the mean difference between two treatment groups includes zero and the mean difference between treatments is not significant, the 95% confidence interval may still provide useful information about the direction and size of the effect of the outcome.

Confidence intervals and P values both provide a dichotomous test of the significance of statistical hypotheses based on the critical level of significance. However, confidence intervals are preferable because they enable a direct statement to be made about the size and direction of the difference between treatment groups. However, it is useful for both confidence intervals and P values to be reported in scientific articles because they provide statistical measures that complement each other.

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. What is a P value? *BMJ* 2012;345:e7767.
3 Sedgwick P. Superiority trials. *BMJ* 2011;342:d2981.
4 Sedgwick P. Confidence intervals: predicting uncertainty. *BMJ* 2012;344:e3147.
5 Sedgwick P. What is sampling error? *BMJ* 2012;344:e4285.
6 Sedgwick P. Confidence intervals and statistical significance: rules of thumb. *BMJ* 2012;345:e4960.

Cite this as: *BMJ* 2013;346:f3212

© BMJ Publishing Group Ltd 2013

BMJ: first published as 10.1136/bmj.f3212 on 17 May 2013. Downloaded from http://www.bmj.com/ on 16 October 2023 at CzechELib project (Charles University). Protected by copyright.

For personal use only: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe