

## ENDGAMES

## STATISTICAL QUESTION

**Confidence intervals and statistical significance: rules of thumb**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 compared the effectiveness of cryotherapy with that of salicylic acid for treating plantar warts. A randomised controlled trial was performed. Participants were eligible if aged 12 years or over. Those randomised to cryotherapy had liquid nitrogen delivered by a healthcare professional, with a maximum of four treatments, 2-3 weeks apart. Participants randomised to 50% salicylic acid (Verrugon) treated themselves daily for a maximum of eight weeks.<sup>1</sup>

The primary outcome was complete clearance of all plantar warts at 12 weeks, as confirmed by inspection of digital photographs by two assessors who were blind to treatment allocation. Secondary outcomes included patients' self reported number of days until clearance of plantar warts in the six months after randomisation. The proportion of participants with complete clearance of plantar warts at 12 weeks was higher in the salicylic group (17 of 119 (14.29%) versus 15 of 110 (13.64%); difference 0.65% (95% confidence interval -8.33 to 9.63)). The hazard ratio for self reported time to clearance of plantar warts in the six months after randomisation when salicylic acid was compared with cryotherapy was 0.8 (0.51 to 1.25).

**Questions**

Which of the following statements, if any, can be concluded?

- The percentage difference between treatment groups in complete clearance of plantar warts at 12 weeks was not significant at the 5% level because the associated 95% confidence interval straddled zero.
- The hazard ratio for self reported time to clearance of plantar warts was significant at the 5% level because the associated 95% confidence interval did not straddle zero.
- The hazard ratio for self reported time to clearance of plantar warts was not significant at the 5% level because the 95% confidence interval straddled unity.

**Answers**

Statements *a* and *c* can be concluded, whereas *b* cannot.

A 95% confidence interval is an interval estimate of the population parameter and represents the uncertainty of the sample in estimating the population parameter as a result of sampling error. Confidence intervals have been described in a previous question.<sup>2</sup> There is a unique relationship between the 95% confidence interval and 5% level of significance when hypothesis testing, described below.

At 12 weeks, the proportion of participants with complete clearance of plantar warts was higher in the salicylic acid group (17/119 (14.29%) versus 15/110 (13.64%)). The difference between treatment groups (salicylic acid minus cryotherapy) was 0.65%, with a 95% confidence interval for the population parameter of -8.33% to 9.63%. The null hypothesis states that the proportion of participants with complete clearance at 12 weeks was the same for the salicylic and cryotherapy treatments in the population from which the sample was taken—that is, the difference between treatments in percentage clearance was zero. The alternative hypothesis states that the difference is not zero—that is, the percentage of complete clearance for salicylic acid was less or greater than that for cryotherapy. Because the 95% confidence interval for the difference in percentage clearance straddled zero, it can be inferred that the difference in percentage clearance between the treatment groups was not significant at the 5% level (*a* is true). The researchers reported that the P value for the statistical test of the difference in percentage clearance was 0.89, with the conclusion that there was no evidence of a difference between treatments in the proportion of participants with complete clearance of plantar warts at 12 weeks.

Generally, if the 95% confidence interval for the difference in an outcome variable between two treatment groups straddles zero then the test of the statistical hypotheses for the difference will not be significant at the 5% level. If the 95% confidence interval excludes zero then the test of the statistical hypotheses will be significant at the 5% level, and the null hypothesis will be rejected in favour of the alternative.

The hazard ratio for the self reported time to clearance of plantar warts in the comparison of salicylic acid with cryotherapy was

0.8 (95% confidence interval 0.51 to 1.25). The hazard ratio, also known as a relative hazard, is a ratio of two rates: in the study above it compares the rate of clearance in the salicylic acid group with the rate in the cryotherapy group during the follow-up period. At any time during follow-up, participants in the salicylic acid group were 0.8 times as likely—that is, 20% less likely—to self report clearance of plantar warts when compared with participants in the cryotherapy group. Hazard ratios have been described in a previous question.<sup>3</sup>

The null hypothesis for the statistical test of the hazard ratio states that there was no difference between salicylic acid and cryotherapy in the participants' self reported time to clearance of plantar warts—that is, the hazard ratio was equal to 1 in the population from which the sample was taken. The alternative hypothesis is two sided: the hazard ratio is less than or greater than 1. Because the 95% confidence interval for the hazard ratio of self reported clearance straddled unity (that is, 1) then it can be inferred that the hazard ratio comparing treatments in self reported clearance time was not significant at the 5% level (*c* is true). The researchers reported that the P value for the statistical test of the hazard ratio was 0.33, with the conclusion that there was no evidence of a difference between treatments

in participants' self reported time to clearance of plantar warts during the six months of follow-up.

More generally, if the 95% confidence interval for a ratio comparing treatment groups in an outcome variable—such as a hazard ratio, relative risk, or odds ratio—straddles unity then the test of the statistical hypotheses for the ratio will not be significant at the 5% level. If the 95% confidence interval excludes unity then the test of the statistical hypotheses will be significant at the 5% level, and the null hypothesis is rejected in favour of the alternative. The 95% confidence interval for a ratio will never straddle zero, the lower limit being above zero and the upper limit bounded by infinity (*b* is false).

Competing interests: None declared.

- 1 Cockayne S, Hewitt C, Hicks K, Jayakody S, Kang'ombe AR, Stamuli E, et al. Cryotherapy versus salicylic acid for the treatment of plantar warts (verrucae): a randomised controlled trial. *BMJ* 2011;342:d3271.
- 2 Sedgwick P. Confidence intervals: predicting uncertainty. *BMJ* 2012;344:e3147
- 3 Sedgwick P. Hazard ratios. *BMJ* 2011;343:d5918.

Cite this as: *BMJ* 2012;345:e4960

© BMJ Publishing Group Ltd 2012