

ENDGAMES

STATISTICAL QUESTION

Statistical tests for independent groups: time to event data

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 assessed the effectiveness of low dose, high frequency ultrasound when combined with standard care in healing venous leg ulcers that had previously been hard to heal. A multicentre randomised controlled trial with a study period of 12 months was performed. Ultrasonography was given for 12 weeks, after which participants returned to standard care alone. Control treatment was standard care alone.[1]

Trial participants were recruited if they had at least one venous leg ulcer of more than six months' duration or greater than 5 cm² in area and an ankle brachial pressure index >0.8. In total, 168 people were randomised to ultrasound and standard care and 169 to standard care alone. The primary outcome was time from randomisation until healing of the largest eligible leg ulcer, referred to as the reference ulcer. The researchers reported that no significant difference existed between treatment groups in the time until healing of the reference leg ulcer (P=0.61).

Which one of the following statistical tests would have been used to compare treatment groups in the time until healing of the reference leg ulcer?

- χ^2 test
- Fisher's exact test
- Log rank test
- Logistic regression

Answers

The log rank test (answer c) would have been used to compare treatment groups in the time until healing of the reference leg ulcer.

The study investigated whether the time taken for the reference leg ulcer to heal differed between treatment groups. The time from allocation to treatment until complete healing is referred to as time to event or survival data, described in a previous question.[2]

The log rank test (answer c) was used to test whether there was any difference between the two treatment groups in their survival times across the 12 month study period. As with traditional hypothesis testing, the log rank test starts at the position of

equipoise. The null hypothesis states that there is no difference in time to healing of the reference leg ulcer between treatment groups across 12 months in the population from which the sample was obtained. The alternative hypothesis states that a difference exists: that in the population from which the sample was obtained the survival times for the two treatments were not equal in the first 12 months. The alternative hypothesis is two sided: the survival times in the ultrasound treatment group could be shorter or longer than those in the control group. The reported P value was 0.61, and therefore the null hypothesis was not rejected in favour of the alternative at the 5% critical level of significance. It was concluded that there was no significant difference between treatments in time to healing of the reference leg ulcer.

The log rank test is purely a test of significance, providing a general test of difference in survival times between treatment groups across the study period. It does not provide an estimate of the difference between treatment groups in their survival times. Furthermore, it is not possible to adjust for the effects of confounding when using the log rank test to compare the survival times of two groups. The median survival time and hazard ratio, described in previous questions,[3] [4] could be used to provide estimates of the magnitude of difference between the treatment groups in their survival times.

The χ^2 test (answer a) and Fisher's exact test (answer b) have a similar objective in that they both establish whether the distribution of a variable with two or more categories is equivalent across two or more independent groups. The tests, described in a previous question,[5] are used to compare two or more groups at one time point. Certain conditions must be met for the χ^2 test to be valid. If the χ^2 test is not valid, then Fisher's exact test is used. Regardless of the validity of the χ^2 test, Fisher's exact test can always be used as an alternative to the χ^2 test.

In the above example, the secondary outcomes included the proportion of participants with all ulcers healed by 12 months. At 12 months after randomisation, 72 of the 168 participants allocated to ultrasound and 78 of the 169 in standard care had

all their ulcers healed. Either the χ^2 test or Fisher's exact test could have been used to compare the treatments in the proportion of participants with all ulcers healed at the end of the study. It was reported that there was no significant difference between the groups at 12 months ($P=0.39$ (Fisher's exact test)).

Logistic regression (answer *d*) is used to derive unadjusted and adjusted odds ratios. Described in a previous question,[6] odds ratios describe the likelihood of a disease or condition if a risk factor is present relative to the risk factor being absent.

Competing interests: None declared.

- 1 Watson JM, Kang'ombe AR, Soares MO, Chuang LH, Worthy G, Bland JM, et al; on behalf of the VenUS III Team. Use of weekly, low dose, high frequency ultrasound for hard to heal venous leg ulcers: the VenUS III randomised controlled trial. *BMJ* 2011;342:d1092.
- 2 Sedgwick P. Survival (time to event) data: censored observations. *BMJ* 2011;343:d4816.
- 3 Sedgwick P, Joeekes K. Survival (time to event) data: median survival times. *BMJ* 2011;343:d4890
- 4 Sedgwick P. Hazard ratios. *BMJ* 2011;343:d5918.
- 5 Sedgwick P. Statistical tests for independent groups: categorical data. *BMJ* 2012;344:e344.
- 6 Sedgwick P. Odds ratios and adjusting for confounding. *BMJ* 2010;341:d6833.

Cite this as: [BMJ 2012;345:e5257](https://doi.org/10.1136/bmj.e5257)

© BMJ Publishing Group Ltd 2012