

ENDGAMES

STATISTICAL QUESTION

Statistical tests for independent groups: categorical data

Philip Sedgwick *senior lecturer in medical statistics*

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

Researchers carried out a randomised controlled trial to compare the effectiveness of cryotherapy with that of salicylic acid for treating plantar warts. Participants randomised to cryotherapy were treated with liquid nitrogen by a healthcare professional, with a maximum of four treatments, each two to three weeks apart. Participants randomised to 50% salicylic acid (Verrugon) treated themselves daily for a maximum of eight weeks. The primary outcome was complete clearance of plantar warts at 12 weeks.¹

The percentage of participants with complete clearance of plantar warts at 12 weeks was slightly higher in the salicylic acid group (17/119 (14.3%) v 15/110 (13.6%)), although the difference was not statistically significant (P=0.89).

Which one of the following statistical tests would most likely have been used to compare the treatment groups with regard to the percentage of participants with complete clearance of plantar warts at 12 weeks?

- The χ^2 test
- Fisher's exact test
- McNemar's test

Answers

The χ^2 test (answer *a*) would most likely have been used to compare the treatment groups with regard to the percentage of participants with complete clearance of plantar warts at 12 weeks.

The χ^2 test is used to compare two independent groups in the proportion that possesses a particular characteristic. For the test to be valid, criteria described below must be met. The numbers of participants in each treatment group with and without complete clearance are shown in the contingency table². The marginal totals are also shown, including the total sample size for each treatment group, plus total numbers with and without complete clearance.

For the above statistical test, the null hypothesis states that in the population the sample was taken from, there would be no difference between the two treatments in the percentage of

patients with complete clearance of plantar warts at 12 weeks. The P value for the test was P=0.89. Therefore, the null hypothesis was not rejected in favour of the alternative, and it was concluded that there was not a statistically significant difference between treatments in effectiveness.

The test of the statistical hypotheses was undertaken by comparing observed and expected frequencies. The observed frequencies—the actual numbers of participants in each treatment group with and without complete clearance—are shown in the contingency table. The expected frequencies are those that would have been obtained if the null hypothesis was true. If the null hypothesis was true and there was no difference between treatments, the percentage of participants with complete clearance in each treatment group could be estimated by the overall percentage for the marginal total with complete clearance. Overall, 32 of 229 (14.0%) participants had complete clearance, so 14.0% of each treatment group would be expected to have complete clearance if the null hypothesis was true. Therefore, the expected numbers of participants achieving complete clearance would be $(32/229) \times 119 = 16.6$ for salicylic acid and $(32/229) \times 110 = 15.4$ for cryotherapy. The expected numbers of participants not achieving complete clearance are derived similarly. Overall, 197 of 229 (86.0%) of participants did not achieve complete clearance. Therefore, the expected number of participants without complete clearance would be $(197/229) \times 119 = 102.4$ for salicylic acid and $(197/229) \times 110 = 94.6$ for cryotherapy. Expected frequencies rarely take integer values, but will always sum to the row and column marginal totals. The P value was derived by comparing the observed and expected frequencies using a formula that can be found in any standard statistical text. If the observed and expected frequencies were equal, there would have been no evidence to reject the null hypothesis in favour of the alternative. As differences between observed and expected frequencies become greater, the evidence against the null hypothesis increases, and it will eventually be rejected in favour of the alternative hypothesis.

The contingency table shown is referred to as a 2x2 table with four cells, there being two treatment independent groups with

each participant having one of two possible outcomes. More generally, the χ^2 test is used to test if the distribution of a variable with two or more categories is equivalent across two or more independent groups. Certain conditions must be met for this test to be valid; no more than 20% of the cells can have an expected frequency less than five and all of the cells must have an expected frequency greater than one. In this example, all of the expected frequencies were larger than five so the χ^2 was valid. If the χ^2 test had been invalid, then Fisher's exact test would have been used (answer *b*). Regardless of the validity of the χ^2 test, Fisher's exact test can always be used as an alternative. However, when the χ^2 test is valid it would usually be used because the calculations for Fisher's exact test are tedious. Performing Fisher's exact test can be computationally exhaustive and can impede the performance of the statistical software, especially when the sample size is large or the contingency table is larger than 2x2.

McNemar's test (answer *c*), described in a previous question,² is used to compare two groups that are related or dependent. Each participant is measured on two occasions in an outcome variable that is dichotomous. The purpose of the test is to establish the extent of agreement between paired measurements across sample members.

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. Statistical tests: matched pairs categorical data. *BMJ* 2011;343:d6623.

Cite this as: *BMJ* 2012;344:e344

© BMJ Publishing Group Ltd 2012

Table

Table 1| Comparison of treatments of cryotherapy and salicylic acid in complete clearance of plantar warts at 12 weeks

Complete clearance at 12 weeks	Treatment group		Total
	Salicylic acid	Cryotherapy	
Yes	17 (14.3%)	15 (13.6%)	32 (14.0%)
No	102 (85.7%)	95 (86.3%)	197 (86.0%)
Total	119 (100%)	110 (100%)	229 (100%)

Percentages shown are those within the treatment group or the overall total.