

ENDGAMES

STATISTICAL QUESTION

Non-parametric statistical tests for two independent groups: numerical data

Philip Sedgwick *reader in medical statistics and medical education*

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

Researchers described the outcomes at one year for a national cohort of infants with gastroschisis. A prospective cohort study design was used. Participants were 301 liveborn infants with gastroschisis between October 2006 and March 2008 from all 28 paediatric surgical centres in the United Kingdom and Ireland. The aim of the study was to describe outcomes at one year, comparing infants with simple gastroschisis (intact, uncompromised, and continuous bowel) with those with complex gastroschisis (bowel perforation, necrosis, or atresia). The main outcome measures included duration of parenteral nutrition and length of stay in hospital.¹

The duration of parenteral nutrition and length of stay in hospital did not follow a normal distribution. Therefore, the groups of infants were compared using non-parametric statistical tests. Infants with complex gastroschisis had a significantly longer duration of parenteral nutrition than those with simple gastroschisis (median 51 days (interquartile range 29-92) v 23 days (16-38); $P < 0.001$). Those with complex gastroschisis also needed a significantly longer stay in hospital (median 84 days (47-197) v 36 days (23-57); $P < 0.001$).

The researchers concluded that the national cohort provided a benchmark against which individual centres could measure outcome and performance. The stratification of neonates with gastroschisis into simple and complex groups reliably predicted outcome at one year.

Which of the following statistical tests could have been used to compare the groups of infants with simple and complex gastroschisis in the duration of parenteral nutrition?

- Kruskal-Wallis test
- Mann-Whitney U test
- Wilcoxon rank sum test
- Wilcoxon signed ranks test

Answers

The Kruskal-Wallis test (answer *a*), Mann-Whitney U test (answer *b*), and the Wilcoxon rank sum test (answer *c*) could all have been used to compare the duration of parenteral nutrition

in infants with simple gastroschisis versus those with complex gastroschisis.

A previous question described how two types of statistical methods—parametric and non-parametric tests—are used to undertake statistical hypothesis testing.² Parametric tests make the assumption that the variable being analysed has a particular distribution—often a normal one—in the population. The normal distribution has been described in a previous question.³ Non-parametric methods make no assumptions about the distribution of the data in the population and are sometimes referred to as distribution-free methods or methods of rank order.

In the above study, the duration of parenteral nutrition did not follow a normal distribution. Therefore, the duration of parenteral nutrition in the groups of infants with simple and complex gastroschisis were compared by using a non-parametric statistical test. The two groups of infants were independent of each other—that is, infants could belong to only one group. The Mann-Whitney U test (answer *b*) and Wilcoxon rank sum test (answer *c*) are non-parametric tests that compare two independent groups in a variable measured on a continuous or ordinal scale. The tests give the same P value and therefore result in the same conclusion with respect to statistical hypothesis testing. The null hypothesis stated that the distribution of the duration of parenteral nutrition for the groups of infants with simple and complex gastroschisis was the same in the population—that is, the median duration of parenteral nutrition for the two groups in the population was equal. Statistical testing resulted in a P value < 0.001 . Therefore, the null hypothesis was rejected in favour of the alternative—that in the population the two groups of infants differed in their distribution of the duration of parenteral feeding. Those infants with complex gastroschisis had a significantly longer duration of parenteral nutrition than those with simple gastroschisis (median 51 v 23 days).

The Kruskal-Wallis test (answer *a*) is a non-parametric test and an extension of the Mann-Whitney U and Wilcoxon rank sum tests. The Mann-Whitney U and Wilcoxon rank sum tests compare the distribution of a variable between two independent

groups, whereas the Kruskal-Wallis test is used to compare the distribution of a variable between three or more independent groups. However, most statistical software packages permit the Kruskal-Wallis test to be used when there are only two independent groups, and in such circumstances the resulting P value is equivalent to that obtained using the Mann-Whitney and Wilcoxon rank sum tests. Therefore, the same conclusion with respect to statistical hypothesis testing would be made.

In the above example, the duration of parenteral nutrition did not have a normal distribution. The distribution of the duration of parenteral nutrition was skewed, and therefore a non-parametric test was used to compare the groups of infants with simple and complex gastroschisis. Although not discussed by the researchers, the assumptions of parametric statistical testing may have been met after transformation of the data. A previous question described how a log transformation of the observations in the variable being tested is often used to achieve a normal distribution.⁴ If the variable being compared between the two groups has a normal distribution, then the Student's *t* test may be used to undertake statistical hypothesis testing.⁵ The Wilcoxon signed ranks test (answer *d*) is a non-parametric test that compares two related groups in a variable that is

continuous or ordinal. The participants in the two samples must be matched or paired. The Wilcoxon signed ranks test has been described in a previous question.⁶ If the samples were paired then, for example, each participant would be measured on two occasions before and after an intervention. If the two groups were matched then there would be a pair of participants, with one in each group, matched on a series of variables such as age and sex.

Competing interests: None declared.

- 1 Bradnock TJ, Marven S, Owen A, Johnson P, Kurinczuk JJ, Spark P, et al; on behalf of BAPS-CASS. Gastroschisis: one year outcomes from national cohort study. *BMJ* 2011;343:d6749.
- 2 Sedgwick P. Parametric v non-parametric statistical tests. *BMJ* 2012;344:e1753.
- 3 Sedgwick P. The Normal distribution. *BMJ* 2010;341:c6085.
- 4 Sedgwick P. Log transformation of data. *BMJ* 2012;345:e6727.
- 5 Sedgwick P. Independent samples *t* test. *BMJ* 2010;340:c2673.
- 6 Sedgwick P. Non-parametric statistical tests for two related groups: numerical data. *BMJ* 2012;344:e2537.

Cite this as: *BMJ* 2014;348:g2907

© BMJ Publishing Group Ltd 2014