AN INTRODUCTION TO STATISTICS

Article 8. An introduction to hypothesis testing. Non-parametric comparison of two groups—1

P Driscoll, F Lecky

Objectives

- Dealing with unpaired non-parametric data
- Dealing with small samples of nominal data

In covering these objectives the following terms will be introduced:

- χ^2 test
- Fisher's exact test
- Yates's correction

In the previous article the basic principles behind comparing two groups was discussed. The t test was also shown to play an important part in this process when dealing with parametric data. As it is a versatile test, it can be used to compare independent groups and paired groups as well as give an estimate of a population mean when a sample is known. In contrast, when dealing with non-parametric data, several tests have to be used. Fortunately though the same principles apply.

As these non-parametric tests do not assume the distribution of the data is normal, they can be used on a much wider spectrum of results. The cost of this is they lack power and are mainly tests of significance. Consequently they will tell if a difference exists but not how big it is (table 1).

There are many non-parametric tests and choosing the most appropriate one can be difficult. Studying published papers can add to the confusion because tests are selected for a variety of reasons, including personal preferences. Furthermore, calculations are commonly carried out with the aid of computer software. As a result there is a risk that the unwary can produce a figure with a p value that is in fact meaningless because the wrong test has been used.

For those unfamiliar with statistics the way forward is to talk to someone who knows about the subject. To make these meetings more useful, the next two articles will concentrate on the common non-parametric tests used for two groups comparisons (table 2).

Table 2 Table of non-parametric tests for two group comparisons

Nominal data	Ordinal data
Unpaired: χ^2 Fisher exact test	Ranked data: Paired—Wilcoxon Unpaired—Mann-Whitney
Paired: McNemar's test	

 χ^2 test**OVERVIEW**

In the medical literature there are many examples of studies that have used nominal data. As described in article 1, this is where the data are divided into categories that do not have any inherent order.¹ It is possible to see how the proportions of observations in different categories from a sample compare ("fit") with that found in a population. Consequently this is known as "A goodness of fit" test. It is also possible to compare the distribution between samples as well as to test the association between variables. In all cases the systematic approach described already in this series is used (box 1).² However, the z and t test is replaced with the χ^2 test.

Goodness of fit

To demonstrate this consider the following example. Dr Egbert Everard, SpR in Emergency

Box 1 System for statistical comparison of two groups

- State the null hypothesis and the alternative hypothesis of the study
- Select the level of significance
- Establish the critical values
- Select the groups
- Choose and calculate the test statistic
- Compare the calculated test statistic with the critical values
- Express the chances of obtaining results at least this extreme if the null hypothesis is true

Key point 1

The χ^2 test is a non-parametric test of the null hypothesis. It can be used on unpaired nominal data to determine:

- A goodness of fit between a sample and a population
- Comparison between two groups
- Association between two variables

Accident and
Emergency
Department, Hope
Hospital, Salford
M6 8HD, UK

Correspondence to:
Mr Driscoll, Consultant in
Accident and Emergency
Medicine
(pdriscoll@
hope.srht.nwest.nhs.uk)

Table 1 Contrast of parametric and non-parametric tests in two group comparisons

	Parametric	Non-parametric
Assumptions	Normal distribution Uniform variance	Depends on test
Data	Paired and unpaired data	Depends on test
Significance	Assessed	Assessed
Small differences	Detected	Missed
Size of the difference	Can be calculated	Cannot be calculated

Medicine, is asked to determine if the type of complaints received by the Emergency Department of Deathstar General is similar to the national picture (table 3). To answer this he decides to use a χ^2 “goodness of fit” test.

Table 3 Complaints made by patients attending UK emergency departments

	Misdiagnosis	Waiting times	Communication	Other
National proportions	15%	60%	10%	15%

Figure 1 Graph of χ^2 values (horizontal axis) against a measure of probability (vertical axis). The curves are theoretical χ^2 distributions for 5, 10 and 15 degrees of freedom (df). Each are skewed to the right, particularly when the df is small. As the df increases the curve becomes more normal in distribution. The total area under each curve is equal to 1 (that is, full probability).

	<i>p</i>			
	0.1	0.05	0.01	0.001
df				
1	2.706	3.841	6.635	10.827
2	4.605	5.991	9.210	13.815
3	6.251	7.815	11.345	16.268
4	7.779	9.488	13.277	18.465
5	9.226	11.070	15.086	20.517

Figure 2 Part of the χ^2 table showing probability (*p*) for various degrees of freedom (df).

1 STATE THE NULL HYPOTHESIS AND THE ALTERNATIVE HYPOTHESIS OF THE STUDY
Having considered the problem, Egbert writes down the null hypothesis as:

“There is no significant difference between the type of complaints made by patients attending the Emergency Department of Deathstar General and those attending emergency departments nationally”.

The alternative hypothesis is the logical opposite of this—that is:

“There is a significant difference between the complaints made by patients attending the Emergency Department of Deathstar General and those attending emergency departments nationally”.

2 SELECT THE LEVEL OF SIGNIFICANCE

If the null hypothesis is correct, the distribution of complaints found in the Emergency Department of Deathstar General should be the same as that found nationally. Therefore the difference between them would be zero. However, there is bound to be some small variation simply due to chance. The question is how big a difference are we going to allow before we reject the idea that the two are all part of the same population?

We can answer this because the difference between a sample and population of nominal data has a χ^2 distribution (fig 1). Egbert picks a significance level of 0.05 because, by convention, the outer 5% of the area under the curve is considered to be sufficiently away from the population mean as to represent values that cannot be attributed to chance variation. He now needs to determine the critical value for the χ^2 statistic that demarcates this point (χ^2_{crit}).

3 ESTABLISH THE CRITICAL VALUES

As with the *t* distribution, the χ^2 distribution changes shape depending upon the size of the sample. For mathematical reasons this is measured as the degrees of freedom rather than the actual size. When comparing a sample with a population, the degrees of freedom is one less than the number of categories in the sample. Therefore in this case, Egbert works out the degrees of freedom to be $4 - 1 = 3$.

Using the table of χ^2 statistics, a value of 7.82 or greater for χ^2_{crit} demarcates a right tail that has, at most, 5% of the area under the curve (fig 2). In other words, 7.82 separates the left 95% area of acceptance of the null hypothesis from the 5% area of rejection.

A sample of patients from Deathstar’s Emergency Department can now be selected and the experimental χ^2 statistic calculated (χ^2_{calc}).

4 SELECT THE SAMPLE

Using records for the previous year, Egbert finds out that the type of complaints made (table 4).

5 CHOOSE AND CALCULATE THE TEST STATISTIC

If the null hypothesis applies there is no difference between the national picture and Deathstar General. Consequently the expected number of complaints in each category (E) should be the same as those actually observed

Table 4 Observed complaints made by patients attending Deathstar's Emergency Departments

	Misdiagnosis	Waiting times	Communi-cation	Other	Total
Observed number	25	15	10	10	60

(O). The value for E in each category can be determined from the percentages provided for the population's categories. For example, in the national picture 15% of complaints are attributable to a misdiagnosis being made. Therefore if the null hypothesis is applied, 15% of patients attending the Emergency Department of Deathstar General should also make this type of complaint. The expected number would therefore be:

$$[15 \times 60]/100 = 9$$

Carrying out the same process for each of the other categories, Egbert draws up a table of expected numbers of complaints if the null hypothesis applied (table 5).

Table 5 Expected complaints from patients attending Deathstar's Emergency Departments

	Misdiagnosis	Waiting times	Communi-cation	Other	Total
Expected number	9	36	6	9	60

To be able to use the χ^2 test, the data need to have the following properties:

- Only frequency data are used in the categories
- Events are independent within a sample group. This means that paired data cannot be used and individual subjects only appear once in the table.
- No expected value in the table is less than 1 and 80% have values over 5. The reason for this is small expected values can have a disproportionate effect on the overall χ^2 test statistic, irrespective of the values in other cells
- There is a logical basis for the group classification.

As these assumptions are valid in this study, Egbert can proceed with determining χ^2_{calc} .

An indicator of the validity of the null hypothesis would be the total differences between the observed (O) and expected (E) values in each category. However, if we simply added up each of these values, the overall result would be zero. This is because half the differences will be positive and the other half negative. To overcome this they are squared first. The χ^2 statistic for each category is then derived by dividing this figure by E. Therefore $(O-E)^2/E$ represents the test statistic for each category. The overall test statistic (χ^2_{calc}) is the sum of these separate category test statistics. This can be represented by the equation:

$$\chi^2_{\text{calc}} = \sum [(O-E)^2 / E]$$

Egbert therefore determines the χ^2 value for each category and derives the overall χ^2 statistic (table 6).

$$\chi^2_{\text{calc}} = \sum [(O-E)^2 / E] = 43.43$$

Key point 2

The order of the categories has no effect on the value of χ^2 . Only the size of the differences between the categories is important.

Table 6 Deriving the overall χ^2 statistic for the study

	Misdiagnosis	Waiting time	Communi-cation	Other
Observed (O)	25	15	10	10
Expected (E)	9	36	6	9
O - E	16	-21	4	1
(O - E) ²	256	441	16	1
(O - E) ² /E	28.4	12.25	2.67	0.11

When carrying out this calculation by hand there are two checks to ensure a simple mistake has not been made:

- The sum of the expected frequencies for all the categories should equal the total observed frequencies for all the categories (that is, 60 in this case).
- In each column, the sum of all observed frequencies minus the sum of all expected frequencies must equal zero.

6 COMPARE THE CALCULATED TEST STATISTIC WITH THE CRITICAL VALUE

The calculated value of 43.43 lies beyond the critical value of 7.82. It therefore falls outside the area of accepting the null hypothesis.

7 EXPRESS THE CHANCES THAT THE NULL HYPOTHESIS IS IN KEEPING WITH THE DATA

As described previously, the p value is the probability of getting a difference equal or greater than that found in the study (that is, 43.43) if the null hypothesis was correct.²

In contrast with the *t* and *z* statistics, only the right side of the χ distribution is used. This is because only large values of χ^2 can reject the null hypothesis.

Key point 3

In other words χ^2 tests are always one sided.³

From the χ^2 tables, it can be seen that the size of the tail from 43.43 to the right tip is less than 0.001. In other words, there is less than a 0.1% chance that a difference with a magnitude of 43.43, or larger, could have resulted if the null hypothesis was valid. Therefore the p value is <0.001.

When presenting the results it is important that they are interpreted in the light of the data. Consequently Egbert concludes that there is a significant difference between the types of complaints made by Deathstar's Emergency Department attendees and those found nationally, $\chi^2 = 43.43$, df 3, $p < 0.001$. The difference is most marked in the complaints made regarding misdiagnoses (greater than expected) and waiting times (less than expected).

Box 2 Summary for calculating the p value using the χ^2 test

- Record the observed category frequencies from the data (O)
- Calculate the expected values (E) for all category frequencies if the null hypothesis was true by $E = \text{row total} \times \text{column total} / \text{grand total}$
- For each cell calculate $(O-E)^2/E$
- Add these values to obtain the test statistic χ^2 where $(\chi^2 = \Sigma(O-E)^2/E)$
- Using tables of the χ^2 distribution, determine the p value for the null hypothesis using the test statistic value and appropriate degree of freedom

Key point 4

When presenting the results of a χ^2 analysis, the χ^2 value, degree of freedom and p value should all be given along with an interpretation in the light of the data.

Comparing the distribution between two groups

A frequent application of the χ^2 test in published work is comparing the distribution of proportions in two groups. To see how this works, consider the following example. Endora Lonely, an Emergency Physician at St Heart-sinc is concerned that Egbert is not getting out enough. She therefore invites him to a meal at her flat. During the evening he talks about his interesting study regarding the emergency department's complaints. He wonders if they are similar to those at St Heartsinc. Amazed by this request she cancels the planned weekend away and, reluctantly, agrees to help the proposed study. As they will be comparing two unpaired groups of nominal data, they decide to use the χ^2 test.

1 STATE THE NULL HYPOTHESIS AND THE ALTERNATIVE HYPOTHESIS OF THE STUDY

Having considered the problem, they write down the null hypothesis as:

"There is no significant difference between the type of complaints made by patients attending the Emergency Departments of Deathstar General and St Heartsinc".

The alternative hypothesis is the logical opposite of this—that is:

"There is a significant difference between the complaints made by patients attending the Emergency Departments of Deathstar General and St Heartsinc".

2 SELECT THE LEVEL OF SIGNIFICANCE

Following convention they pick a significance level of 0.05. They now need to determine the critical value for the χ^2 statistic that demarcates this point (χ^2_{crit}). As described in the previous example, if the null hypothesis is correct, the distribution of complaints found in the two departments should be the same. However, there is bound to be some difference between them due to random variation but this has a χ^2

distribution. This distribution can therefore be used to find the value of χ^2_{crit} .

3 ESTABLISH THE CRITICAL VALUES

With this type of comparison, the degrees of freedom is $(r-1) \times (c-1)$ where r and c are the number of rows and columns in the contingency table (table 7).

Table 7 Contingency table for the study comparing the complaints made by patients attending Deathstar's and St Heartsinc's Emergency Departments

	Misdiagnosis	Waiting times	Communi-cation	Other	Total
Deathstar General					
St Heartsinc					

In this study, they calculate the degrees of freedom to be $(2-1) \times (4-1) = 3$. The χ^2 statistics tables indicate that this gives a χ^2_{crit} value of 7.82 or greater at the 5% level. In other words, 7.82 separates the left 95% area of acceptance of the null hypothesis from the right 5% area of rejection.

4 SELECT THE SAMPLE

Endora now checks the records from St Heart-sinc to find out the type of complaints made by the emergency department patients (table 8).

Table 8 Observed complaints made by patients attending Deathstar's and St Heartsinc's Emergency Departments

	Misdiagnosis	Waiting times	Communi-cation	Other	Row total
Deathstar General	25	15	10	10	60
St Heartsinc	20	5	10	5	40
Column total	45	20	20	15	100

5 CHOOSE AND CALCULATE THE TEST STATISTIC

If the null hypothesis applies there should not be a difference between the two departments. Consequently the expected number of complaints in each category (E) should be the same as those actually observed (O). The best estimate we have for the expected values comes from combining the values for two groups in each category so that an overall proportion can be calculated.

For example, the overall proportion of complaints attributable to a misdiagnosis is $45/100$ (45%). Therefore if the null hypothesis applied, 45% of patients attending the Emergency Department of Deathstar General would also make this type of complaint. The expected number would therefore be:

$$[45 \times 60]/100 = 27$$

Similarly the expected number of misdiagnosis complaints in St Heartsinc is:

$$[45 \times 40]/100 = 18$$

Carrying out the same process for each of the other categories, they draw up a table of expected numbers of complaints if the null hypothesis applied (table 9). Again these data fulfil the assumptions made in using the χ^2 test.

As the overall χ^2 statistic is:

$$\chi^2_{\text{calc}} = \Sigma [(O-E)^2 / E]$$

Table 9 Expected complaints from patients attending Deathstar's and St Heartsinc's Emergency Departments

	Misdiagnosis	Waiting times	Communication	Other	Total
Deathstar					
General	27	12	12	9	60
St Heartsinc	18	8	8	6	40

$$\chi^2_{\text{calc}} = \Sigma [(25-27)^2/27 + (15-12)^2/12 + (10-12)^2/12 + (10-9)^2/9 + (20-18)^2/18 + (5-8)^2/8 + (10-8)^2/8 + (5-6)^2/6]$$

$$\chi^2_{\text{calc}} = 3.84$$

6 COMPARE THE CALCULATED TEST STATISTIC WITH THE CRITICAL VALUE

This calculated value is less than the critical value of 7.82. It therefore falls within the area of accepting the null hypothesis.

7 EXPRESS THE CHANCES THAT THE NULL HYPOTHESIS IS IN KEEPING WITH THE DATA

From the χ^2 tables, it can be seen that the size of the tail from 3.84 to the right tip is greater than 0.05. Consequently Endora and Egbert conclude that there is no significant difference between the types of complaints made by patients attending the two emergency departments, $\chi^2 = 3.84$, df 4, $p > 0.05$.

Association between variables

In the previous example we have seen how the χ^2 test can be used to determine if the observed numbers in each category of the table differ from those expected if the null hypothesis was valid. The same process can be used to identify an association between the column and row variables. This is particularly useful when dealing with nominal data.

Egbert's next study demonstrates this. Having seen a number of life threatening complications resulting from central line insertion, Egbert is a keen supporter of the femoral approach. To try and convince his colleagues at Deathstar General he carries out a retrospective study to see if there is any association between type of approach and life threatening complications.

1 STATE THE NULL HYPOTHESIS AND THE ALTERNATIVE HYPOTHESIS OF THE STUDY

The null hypothesis for this study is:

"The incidence of life threatening complications following central line insertion is independent of the approach".

Again the alternative hypothesis is the logical opposite of this—that is:

"The incidence of life threatening complications following central line insertion is dependent on the approach".

Table 12 Contingency table showing the expected frequency of life threatening complications for different central line approaches

	Subclavian	Internal Jugular	Femoral	Row total
Life threatening complications present	$[70 \times 30]/150 = 14$	$[40 \times 30]/150 = 8$	$[40 \times 30]/150 = 8$	30
No life threatening complications	$[70 \times 120]/150 = 56$	$[40 \times 120]/150 = 32$	$[40 \times 120]/150 = 32$	120
Column total	70	40	40	150

2 SELECT THE LEVEL OF SIGNIFICANCE

A significance level of 0.05 is chosen. If the null hypothesis was valid, the difference between the approaches would have a χ^2 distribution. This distribution can therefore be used to find the value of χ^2_{crit} that corresponds to a significance level of 0.05.

3 ESTABLISH THE CRITICAL VALUES

With this type of comparison, the degrees of freedom are $(r-1) \times (c-1)$ where r and c are, respectively, the number of rows and columns in the contingency table (table 10).

Table 10 Contingency table for the study comparing central line approach and life threatening complications

	Subclavian	Internal Jugular	Femoral	Total
Life threatening complications present				
No life threatening complications				

In this study, the degrees of freedom are $(2-1) \times (3-1) = 2$. The χ^2 statistics tables indicate that this gives a χ^2_{crit} value of 5.99 or greater at the 5% level.

4 SELECT THE SAMPLE

From the emergency department records, Egbert determines the incidence of life threatening complications resulting from central line insertion in the past year (table 11).

Table 11 Contingency table for the study comparing central line approach and life threatening complications

	Subclavian	Internal Jugular	Femoral	Row total
Life threatening complications present	15	10	5	30
No life threatening complications	55	30	35	120
Column total	70	40	40	150

5 CHOOSE AND CALCULATE THE TEST STATISTIC

The overall test statistic is calculated in the same way as described in the previous examples. Egbert therefore calculates the expected values for each category if the null hypothesis applied (table 12). Again these data fulfil the assumptions made in using the χ^2 test.

As the overall χ^2 statistic is therefore:

$$\chi^2_{\text{calc}} = \Sigma [(1)^2/14 + (2)^2/8 + (-3)^2/8 + (-1)^2/56 + (-2)^2/32 + (3)^2/32]$$

$$\chi^2_{\text{calc}} = 2.12 \text{ (rounded up)}$$

6 COMPARE THE CALCULATED TEST STATISTIC WITH THE CRITICAL VALUE

This calculated value is less than the critical value of 5.99. It therefore falls within the area of accepting the null hypothesis.

7 EXPRESS THE CHANCES THAT THE NULL HYPOTHESIS IS IN KEEPING WITH THE DATA
Egbert concludes that there is no association between life threatening complications and the type of central line approach used, $\chi^2 = 2.12$, df 2, $p > 0.05$.

Dealing with small samples

The χ^2 test statistic complies with the χ^2 distribution provided the expected values are large enough. When dealing with small samples we can no longer make this assumption. A way of dealing with this problem is to merge categories so that the expected number in each is greater than 5. Obviously the type categories combined needs to be logical.

Another way of tackling this problem is to use Fisher's exact test.

FISHER'S EXACT TEST

The study by Rogers *et al* study demonstrates the use of this test. They investigated neurogenic pulmonary oedema in patients with low and high intracranial pressure (ICP).⁴ Part of the study involved determining if the patients also had normal or abnormal $\text{PaO}_2/\text{FiO}_2$ ratios. The results were tabulated (table 13).

Table 13 Relation between ICP and $\text{PaO}_2/\text{FiO}_2$ ratios

	Normal $\text{PaO}_2/\text{FiO}_2$	Abnormal $\text{PaO}_2/\text{FiO}_2$
High ICP	9	1
Low ICP	1	12

The data were not normally distributed and there are frequencies less than 5 in two of the cells. Analysis of the data was therefore carried out using Fisher's exact test. This produced a p value < 0.001 . Consequently there is less than 1 in 1000 chance that a difference this big, or greater, between these groups is attributable to chance. The null hypothesis was therefore rejected and it was concluded that patients with high ICP values are more likely to have abnormal $\text{PaO}_2/\text{FiO}_2$ ratios compared with patients low ICP.

For those who are interested, Bland describes the method of calculating this test statistic.⁵ This is rarely necessary because it is typically worked out using a computer program. More commonly though we will come across this test when reading published articles. In these cases it is important to bear in mind the following points:

- It determines the probability that the null hypothesis is correct—that is, both groups have the same proportion of outcome/conditions.
- It is used when the number in one or more of the categories of the contingency table are less than 5.

Table 14 Contrast of parametric and non-parametric tests in two group comparisons

	Parametric	Non-parametric
Assumptions	Normal distribution Uniform variance	
Data	Paired and unpaired data	
Significance	Assessed	
Small differences	Detected	
Size of the difference	Can be calculated	

- It assumes the data are unpaired and only frequency data (not proportions) are used.

Box 3 Statistical trivia

Ronald Fisher was born in London 1870. He received a BA in astronomy from Cambridge in 1912 but in 1919 went to work at the Rothamsted Agricultural Experiment Station where he worked as a biologist and made many contributions to both statistics and genetics. He is quoted as saying, "To call in the statistician after the experiment is done may be no more than asking him to perform a postmortem examination: he may be able to say what the experiment died of".

YATES'S CORRECTION FOR CONTINUITY

The distribution of χ^2 is continuous, yet nominal data are not. This can give rise to the false rejection of the null hypothesis in 2×2 tables unless adjustments are made. The correction entails the following alteration:

$[(\text{Observed frequency} - \text{expected frequency}) - 0.5]^2 / \text{expected frequency}$

This has the effect of reducing the χ^2 value and, consequently, decreasing the p value.

Though there is no precise rule when the Yates's correction should be applied, Altman recommends it is always used when dealing with 2×2 tables.⁶ In that way it has its biggest effect where the risk of bias is highest.

Key points 5

- In 2×2 tables with small frequencies, Yates's correction can have marked effect on the p value.
- This correction is not needed for larger tables or when the χ^2 does not reach statistical significance.

Summary

There are many types of non-parametric test and several are frequently used in the medical journals. In trying to choose the correct test it is important to bear in mind the type of data you are dealing with and the nature of the question.

The χ^2 test is a versatile and commonly used method for comparing distribution and looking for associations between groups of unpaired nominal data. It does however have its limitations, particularly when dealing with small samples. In these cases Yates's correction may need to be included, or the Fisher's exact test carried out instead, depending upon the data.

Quiz

- 1 Complete table 14.
- 2 Harrison *et al* studied the association between head injuries and five types of facial injury sustained by cyclists.⁷ How many degrees of freedom (df) would there be in calculating the χ^2 statistic?
- 3 Egbert takes time off to coordinate a defibrillation course for personnel from St Heartsinc and Deathstar General. Having dispatched several instructors to hospital

Table 15 Dangerous defibrillation by personnel from St Heartsinc and Deathstar General

	Dangerous	Not dangerous	Row total
St Heartsinc	10	20	30
Deathstar General	8	12	20
Column total	18	32	50

Table 16 The male and female incidence of do not attempt resuscitation (DNAR) orders

	No DNAR	DNAR	Row total
Male	211	32	243
Female	201	26	227
Column total	412	58	470

with third degree burns, he concludes that several of the doctors appear to be dangerous (table 15). Determine if there is a difference between the hospitals using a 5% level for the null hypothesis.

- 4 De Vos *et al* studied the impact of gender on do not attempt resuscitation (DNAR) orders in hospital (table 16).⁸ Determine if there is there an association between these variables using a 5% level for the null hypothesis.
- 5 One for you to try on your own. Stancin *et al* conducted a study on the acute psychosocial impact of paediatric orthopaedic trauma victims.⁹ They wanted to compare patients with (group 1) and without (group 2) an accompanying head injury. There were 80 patients in group 1 and 28 in group 2. To determine compatibility between the groups they studied whether the patient was white or not. The results showed there was 46 white patients in group 1 and 24 in group 2. Is there an association between head injury and race?

Answers

1 See table 1.

2 4

$df = (\text{number of rows} - 1) \times (\text{number of columns} - 1)$

Therefore $df = (5 - 1) \times (2 - 1) = 4$

Table 17 Expected frequencies for each cell

	No DNAR	Expected	DNAR	Expected
Male	211	$243 \times 412 / 470 = 213$	32	$243 \times 58 / 470 = 30$
Female	201	$227 \times 412 / 470 = 199$	26	$227 \times 58 / 470 = 28$
	412	412	58	58

Table 18 The χ^2 values for each cell

	No DNAR $[O - E]^2 / E$	DNAR $[O - E]^2 / E$
Male	$[-2]^2 / 213 = 0.019$	$[2]^2 / 30 = 0.133$
Female	$[2]^2 / 201 = 0.02$	$[-2]^2 / 30 = 0.133$

3 There is no significant difference between the two hospitals; $\chi^2 = 0.23$, $df = 1$, $p > 0.05$.

4 Using the χ^2 test the expected frequencies for each cell need to be calculated (table 17).

For each cell, the $[\text{observed frequency} - \text{expected frequency}]^2 / \text{expected frequency}$ ($[O - E]^2 / E$) now needs to be calculated and added together (table 18):

The test statistic is the sum of all the $([O - E]^2 / E) = 0.305$. The degrees of freedom are $(\text{number of rows} - 1) \times (\text{number of columns} - 1) = 1$. Using the χ^2 distribution tables for this degree of freedom shows that the χ^2 value is well above the 5% probability for the null hypothesis being valid. It was therefore concluded that there was no association between DNAR resuscitation orders and gender.

The authors would like to thank Jim Wardrope for his help and support.

- 1 Driscoll P, Lecky F, Crosby M. An introduction to everyday statistics—1. *J Accid Emerg Med* 2000;17:205–11.
- 2 Driscoll P, Lecky F. An introduction to hypothesis testing. Parametric comparison of two groups—1. *Emerg Med J* 2001;18:124–30.
- 3 Driscoll P, Lecky F. An introduction for hypothesis testing. Parametric comparison of two groups—2. *Emerg Med J* 2001;18:214–21.
- 4 Rogers F, Shackford S, Trevisani G, *et al*. Neurogenic pulmonary oedema in fatal and nonfatal head injuries. *J Trauma* 1995;39:860–6.
- 5 Bland M. The analysis of cross-tabulations using χ^2 distributions. In: *An introduction to medical statistics*. Oxford: Oxford University Press, 1987:241–63.
- 6 Altman D. Comparing groups—categorical data. In: *Practical statistics for medical research*. London: Chapman Hall, 1997:229–76.
- 7 Harrison M, Shepherds J. The circumstances and scope for prevention of maxillofacial injuries in cyclists. *J R Coll Surg Edinb* 1999;44:82–6.
- 8 de Vos R, Koster R, de Haan R. Impact of survival preferences on do-not-attempt-resuscitation orders in a hospital. *Resuscitation* 1998;39:15–21.
- 9 Stancin T, Taylor G, Thompson G, *et al*. Acute psychosocial impact of pediatric orthopedic trauma with and without accompanying brain injuries. *J Trauma* 1998;45:1031–8.

Further reading

Koosis D. A test of distributions. In: *Statistics—a self teaching guide*. 4th ed. New York: Wiley, 1997:209–29.

Stephens L. χ^2 procedures. In: *Theory and problems of beginning statistics*. Schaum's outline series. New York: McGraw-Hill, 1998:249–71.

Swinow T. The χ^2 tests. *Statistics from square one*. London: BMJ, 1983:43–3.