non-genitourinary medicine settings. *J Fam Plann Reprod Health Care* 2007; **33**: 177–182.
14 National Chlamydia Screening Steering Group (NCSSG). *First Steps: Annual Report of the National Chlamydia Screening Programme in England, 2003/2004*. London, UK: Department of Health, 2004.
15 National Chlamydia Screening Steering Group (NCSSG). *Looking Back, Moving Forward: Annual Report for the National Chlamydia Screening Programme in England, 2004/05*. London, UK: Department of Health, 2005.
16 National Chlamydia Screening Steering Group (NCSSG). *New Frontiers: Annual Report of the National Chlamydia Screening Programme in England 2005/2006*. London, UK: Health Protection Agency, 2006.
17 Jones K, Webb A, Mallinson H, Birley H. Outreach health adviser in a community clinic screening programme improves management of genital chlamydia infection. *Sex Transm Infect* 2002; **78**: 101–105.
18 Poulin C, Alary M, Bernier F, Carbonneau D, Boily M, Joly JR. Prevalence of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* among at-risk women, young sex workers, and street youth attending community organizations in Quebec City, Canada. *Sex Transm Dis* 2001; **28**: 437–443.
19 Macleod J, Holloway A, Morcom A, Hobbs FDR, Salisbury C, Patel R, *et al*. Coverage and uptake of systematic postal screening for genital *Chlamydia trachomatis* and prevalence of infection in the United Kingdom general population: cross sectional study. *BMJ* 2005; **330**: 940–942.
20 Novak DP, Edman AC, Jonsson M, Karlson RB. The internet, a simple and convenient tool in *Chlamydia trachomatis* screening of young people. *Euro Surveill* 2003; **8**: 171–176.
21 Sripada S, Logan S, McGillivray S, Mckenzie H, Templeton A, Hamilton M, *et al*. Opportunistic screening for *Chlamydia trachomatis* in men attending three different secondary health care settings. *Sex Transm Infect* 2007; **83**: 282–285.
22 Jones CA, Knaup RC, Hayes M, Stoner BP. Urine screening for gonococcal and chlamydial infections at community-based organizations in a high-morbidity area. *Sex Transm Dis* 2000; **27**: 146–151.
23 Sternberg P, Hubley J. Evaluating men's involvement as a strategy in sexual and reproductive health promotion. *Health Promot Int* 2004; **19**: 389–396.
24 Fenton KA. Screening men for *Chlamydia trachomatis* infection: have we explored the possibilities? *Commun Dis Public Health* 2000; **3**: 86–89.
25 Hart G, Duncan B, Fenton KA. Chlamydia screening and sexual health. *Sex Transm Infect* 2002; **78**: 396–397.

# Testing and quantifying association in binary data

**Pamela Warner**

## Background

Both Chi-square test and logistic regression have been used to explore association in an article in this issue of the Journal.[1] These notes are intended to provide readers with some supplementary explanation and comparison of these statistical methods. [See Box 1 for glossary of terms.]

## What is it?

The Chi-square test can be used to test the null hypothesis (NH) of 'no association' between two categorical variables. When both outcome variable and explanatory variable are binary, there are only four possible combinations of values for outcome/explanatory variable, and hence study sample data can be accumulated in a classic 2 × 2 table of frequency counts. [NB. Chi-square tests can be undertaken for larger tables, but these notes consider only 2 × 2 tables.] For binary data, the research question of interest is typically: "Is there an association between outcome (screening uptake) and an explanatory factor (gender)?", for example, "Does uptake differ between males and females?". That is, the test is equivalent to testing the NH of no difference in proportions (or percentages) of individuals with the outcome.

## When and why is it useful?

The Chi-square test is extremely useful in most 2 × 2 tables for testing association. It examines the *observed* cell counts (the data from the study sample), and compares these to cell counts that would be expected if there were truly no association between the two variables. The Chi-square value calculated from the sample data is referred to tables of the Chi-square distribution, to ascertain the significance probability (under the NH). If this probability is sufficiently low (conventionally <5% or <1%) then the data are judged too unlikely for the NH to be true, so we conclude, by reverse logic, that the NH must be false, that there is therefore an association.

*J Fam Plann Reprod Health Care* 2009; **35**(1): 26–27

**Public Health Sciences, University of Edinburgh Medical School, Edinburgh, UK**
Pamela Warner, BSc, PhD, *Senior Lecturer in Medical Statistics and Associate Editor, Journal of Family Planning and Reproductive Health Care*

**Correspondence to:** Dr Pamela Warner, Public Health Sciences, University of Edinburgh Medical School, Teviot Place, Edinburgh EH8 9AG, UK. E-mail: p.warner@ed.ac.uk

## What precautions are needed?

The validity of the Chi-square test for 2 × 2 tables is good if the total $n$ is greater than 40. If the total $n$ is between 20 and 40 then validity remains good provided none of the four expected cell counts is less than 5. Otherwise, or if total $n$ is less than 20, Fisher Exact test should be used instead of Chi-square.[2]

The Chi-square test does *not* provide a measure of the *degree* of association. The significance probability cannot serve this purpose, since it reflects the overall $n$ as well as the degree of association. Therefore in reporting Chi-square results it is recommended to present, in addition to the numbers and percentages with the outcome in both explanatory variable subgroups, a summary statistic estimating the association, preferably with a confidence interval for the estimate. Possibilities for the summary are the difference in proportions/percentages or, alternatively, a ratio summary statistic [e.g. odds ratio (OR)].

## Example of techniques

Table 1 shows the results when applying (to the data reported by Lorimer *et al*. in Table 1[1]) the Chi-square test of association of screening uptake and sex, both separately by setting and overall (ignoring setting). Across settings (which are fairly similar in size, $n$ = 104 to 127), the more extreme the sex difference in uptake percentage, the larger is the Chi-square value, and smaller (more significant) the $p$ value. The difference *overall* is similar to that for the education setting (about 13 percentage points), but the much larger overall $n$

**Table 1** Association of screening 'uptake' with gender: tests and summaries – overall, separately by setting, and adjusted for setting

| Analysis | Study *n* | Difference in percentage 'uptake' (female-male pp)[a] | Chi-square ($\chi^2$) test | | Logistic regression | | *p* value if study *n* doubled[b] |
|---|---|---|---|---|---|---|---|
| | | | $\chi^2$ value | Significance probability (P) | Odds ratio (OR)[a] | *p* value | |
| Separately by setting | | | | | | | |
| Workplace | 104 | –7.5 pp | 0.7 | 0.407 | 0.69 | 0.408 | 0.2406 |
| Education | 115 | –13.2 pp | 3.1 | 0.078 | 0.42 | 0.084 | 0.0127 |
| Health and fitness | 127 | –23.1 pp | 6.6 | 0.010 | 0.39 | 0.011 | 0.0003 |
| Overall (ignoring setting) | 346 | –13.3 pp | 6.9 | 0.009 | 0.546 | 0.009 | 0.00021 |
| 'Adjusted for' setting[c] | 346 | ~ | ~ | ~ | 0.472 | 0.002 | 0.00002 |

[a]Confidence intervals omitted here for brevity.
[b]Supposing associations unchanged but $n$ = 208, 230 and 254 for the three settings, 692 overall.
[c]By multivariate logistic regression.
pp, percentage points; ~, not possible.

(i.e. 346 vs 115) results in a larger Chi-square value, and hence a much smaller $p$ value (i.e. 0.009 vs 0.078).

Table 1 also presents the ORs and their $p$ values, obtained by means of univariate logistic regression analysis of association between uptake and sex (as reported for 'overall' in the first column of Lorimer *et al*. Table 2).[1] The $p$ values are very similar to those obtained for Chi-square and a similar pattern across settings is seen for ORs, in that the more extreme the difference in percentages, the more extreme the OR (smaller/further from 1).

Multivariate logistic regression allows testing/estimation of the association *adjusted* for setting. The $n$ is unchanged from the overall analysis ($n$ = 346) but the $p$ value is considerably more significant (0.0002 vs 0.009), partly because the OR is now more extreme (0.472 vs 0.546), but mainly because more precise estimation is possible by combining 'within setting' estimates. As reported by Lorimer *et al*. in Table 2,[1] this association is estimated to be even more extreme if adjusted for both setting and age group: OR = 0.42, a 58% lower odds of uptake of screening for females compared to males.

Finally, the analyses have been rerun for the hypothetical situation that the study size is doubled, but the associations within settings and overall are unchanged (i.e. unchanged percentage point differences and ORs). The $p$ values for these are given in the final column of Table 1 above, and it can be seen how much lower they are, *despite identical associations*, simply because they are derived from a larger study. This highlights the fact that the $p$ value is not a dependable indicator of degree of association.

## Overview

The Chi-square test is easy to apply and suitable for testing association in most 2 × 2 tables. Since $p$ values reflect $n$, reporting of results is more informative if the test is supplemented with a summary statistic of the degree of the association (e.g. OR). Univariate logistic regression analysis of a 2 × 2 association will give very similar statistical significance to Chi-square. Logistic regression has the advantage that multivariate analyses are possible, allowing adjustment for other variables that might be affecting the estimate of the association between outcome and the explanatory variable of interest.

### Statements on funding and competing interests

**Funding** None identified.
**Competing interests** None identified.

### References

1 Lorimer K, Reid ME, Hart GJ. Willingness of young men and women to be tested for *Chlamydia trachomatis* in three non-medical settings in Glasgow, UK. *J Fam Plann Reprod Health Care* 2009; **35**: 21–26.
2 Kirkwood BR, Sterne JAC. *Essential Medical Statistics*. Oxford, UK: Blackwell Science, 2003.

**Box 1: Glossary of statistical terms used in this article**

| | |
|---|---|
| **'Adjusted' association** | See *Association* and *Logistic regression*. |
| **Association** | Relationship between two variables. For binary variables this means that the occurrence of a particular value of one variable, in an individual, is associated with (more likely to be in conjunction with) a particular value of the other variable. |
| **Binary variable** | Has only two possible values (e.g. accept screening or not, female or not). |
| **Categorical variable** | Has a set of distinct values such as gender, recruitment setting. |
| **Chi-square test** | A test applied to counts of individuals cross-tabulated by two categorical variables, to assess association between them (i.e. 'non-independence'). |
| **Chi-square value** | The test value, cumulated either using the expected frequencies and differences between observed and expected, or for 2 × 2 tables, an equivalent short-cut formula applied to observed counts. |
| **Expected** (cell counts) | Calculated from total $n$, and the overall numbers of individuals with the outcome and with the explanatory feature. |
| **Explanatory variable** | A feature potentially associated with outcome. |
| **Fisher Exact test** | This calculates by combinatorial algebra the exact probability, if the NH is true, of obtaining the observed cell counts, or any more extreme arrangement. |
| **Logistic regression (LR)** | LR estimates and tests association between a binary outcome and one or more explanatory variables, with association being summarised as ORs. In univariate LR there is only one explanatory variable and, if that is binary, only one OR. If there is more than one explanatory variable then multivariate LR is needed (MV LR). This estimates the association of outcome with each explanatory variable *'adjusted'* for the joint associations with other explanatory variables in the model, by 'averaging' the separate estimates obtained within cross-classified subgroups (in the example used here, just settings). |
| **Null hypothesis (NH)** | A statement, prior to testing, of no effect (in this case, no association). See also *Significance probability.* |
| **Odds ratio (OR)** | The OR is the ratio of the odds of outcome (uptake) in those with the explanatory feature (female) relative to the odds in those *without* it (male). When there is no association, the two odds should be approximately equal and the ratio approximately 1, or 'null '. The more extreme the OR (away from 1, i.e. 0.4 vs 0.7, or 3.2 vs 1.8) the greater the degree of association. OR <1 means inverse association. |
| **Outcome variable** | Indicates status with respect to the condition of interest (e.g. uptake of screening). |
| **Significance probability** | The probability of obtaining the observed cell counts if the NH is true. |