# BIOSTAT Case Study: Tests of Association for Categorical Data

### LEARNING OBJECTIVES

At the completion of this Case Study, participants should be able to:

* Compare two or more proportions
* Calculate and interpret confidence intervals for proportions
* Understand the impact of expected values on the choice of statistical test used to compare proportions
* Interpret the results of tests of association
* Interpret logistic regression results.

This material was developed by the staff at the Global Tuberculosis Institute (GTBI), one of four Regional Training and Medical Consultation Centers funded by the Centers for Disease Control and Prevention. It is published for learning purposes only. Permission to reprint excerpts from other sources was granted.

Case study author(s) name and position:

Marian R. Passannante, PhD

Associate Professor, University of Medicine & Dentistry of New Jersey, New Jersey Medical School and School of Public Health

Epidemiologist, NJMS, GTBI

For further information please contact:

New Jersey Medical School Global Tuberculosis Institute (GTBI)

225 Warren Street P.O. Box 1709  
Newark, NJ 07101-1709

or by phone at 973-972-0979

**Suggested Citation: New Jersey Medical School Global Tuberculosis Institute. /Incorporating Tuberculosis into Public Health Core Curriculum./ 2009: BIOSTATISTICS CASE STUDY 2: Tests of Association for Categorical Data STUDENT Version 1.0.**

**Introduction**

This exercise is based on the following study. Sections of this document have been reprinted with permission of the journal.

**Factors influencing the successful treatment of infectious pulmonary tuberculosis** W-S. Chung,\*† Y-C. Chang,† M-C. Yang†, \* Department of Internal Medicine, Hualien General Hospital, Hualien, † Institute of Health Care Int J Tuberc Lung Dis 11:59–64 © 2007 The Union

The abstract states that “(t)his study used a population-based…design. All PTB [pulmonary TB] patients residing in southern Taiwan recorded in the tuberculosis registry from 1 January to 30 June 2003 were identified. Each patient’s medical record was requested from treating hospitals and retrospectively reviewed for 15 months after the date PTB was confirmed.” 1

Following is the methods section of this article1.

METHODS

We carried out a population-based medical record review in southern Taiwan, where the only chest specialty hospital geared towards specialized thoracic disease care, mainly for TB, is located. Hospitals and primary practitioners that provided TB care in the same region can be used as comparative care providers. Study areas include Chiayi County, Chiayi City, Tainan County and Tainan City. As mandated by law in Taiwan, all suspected and confirmed TB cases must be reported in a timely manner to the national computerized registry maintained by the Taiwan Center for Disease Control (CDC). Reporting of cases has been encouraged and reinforced through the implementation of a no-notification, no-reimbursement policy and a notification-for-fee policy since 1997. 7 We requested data on all suspected and confirmed TB patients residing in the studied areas and recorded in the registry for the period 1 January to 30 June 2003. The study team, including four registered nurses (each with a minimum of 6 years’ clinical experience), two head nurses (each with a minimum of 12 years’ clinical experience) and one pulmonologist, had undergone a series of training courses designed to ensure proper validation of data consistency. Site visits were arranged to review the medical record of each patient, and the 15-month follow-up of medical records after start of treatment was reviewed.

**Health care institutions**

Health care institutions that had ever reported cases in the study areas included the chest hospital, two academic medical centers, 11 regional hospitals and 15 district hospitals and primary practitioners (district hospitals and primary practitioners are regarded as being at the same level in terms of TB treatment). In Taiwan, institutions are classified by the government as follows: ‘medical centers’ are health care, training and research facilities that house over 500 acute-care beds; ‘regional hospitals’ have no fewer than 250 acute care beds and are staffed by physicians of various specialties with the purpose of providing health care services to patients and training for specialists; and ‘district hospitals’ provide primary health care services similar to those offered by primary practitioners but with the added availability of in-patient care.

**Infectious PTB**

Infectious PTB is defined as sputum culture-confirmed disease caused by *Mycobacterium tuberculosis*, or two sputum smear examinations positive for acid-fast bacilli (AFB) or one positive sputum examination, radiological signs and a clinician’s decision to treat.8

**Directly observed treatment**

For directly observed treatment (DOT), a health worker or other trained person who is not a family member watches as the patient swallows anti-tuberculosis medicines for at least the first 2 months of treatment.1 DOT thus shifts the responsibility for cure from the patient to the health care system. In Taiwan, whether or not the patient is receiving DOT, TB is treated using WHO-recommended regimens; the initial phase consists of 2 months of isoniazid (H), ethambutol (E), rifampicin (R) and pyrazinamide (Z), followed by a 4-month continuation phase consisting of H, E and R (2HERZ/4HER).9,10

**Treatment success**

Treatment success is defined as a patient who has been cured or has received a complete course of treatment. A cured case is defined as a PTB patient who has finished treatment with a negative bacteriology result during and at the end of treatment. A case recorded as completed treatment is defined as a PTB patient who has finished treatment, but who has not met the criteria to be defined as a cure or a failure.11,12

**Ethical consideration**

The study was approved by the Taiwan CDC. All staff members involved in the study signed a statement of agreement to maintain patient confidentiality.

**Data analysis**

Bivariate analyses with 2 tests were used to compare differences in proportions of dichotomous and categorical variables, which extracted potential predictors of successful treatment. We then performed multivariate logistic regression analyses on the potential predictors with *P <* 0.10 obtained from bivariate analyses. We constructed a full model that included all the potential predictors identified through bivariate analyses and then applied the forward substitution model building procedure to construct a reduced model in which all the predictors were statistically significant. Odds ratios (ORs) and 95% confidence intervals (CIs) of dichotomous and categorical risk variables on the binary outcome variables were calculated. All analyses were conducted using SPSS 10.0 software (SPSS Inc, Chicago, IL, USA), and all the tests were performed at the two-tailed significance level of 0.05.

References that appear in the excerpt from this article:

1 World Health Organization. Tuberculosis Fact Sheet. Geneva,Switzerland: WHO. [http://www.who.int/mediacentre/factsheets/fs104/en/index.html Accessed August 2006](http://www.who.int/mediacentre/factsheets/fs104/en/index.html%20Accessed%20August%202006).

7 Chiang C Y, Enarson D A, Yang S L, Suo J, Lin T P. The impactof National Health Insurance on the notification of tuberculosis in Taiwan. Int J Tuberc Lung Dis 2002; 6: 974–979.

8 Migliori G B, Raviglione M C, Schaberg T, et al. Tuberculosis management in Europe. Task Force of the European Respiratory Society, the World Health Organization and the International Union Against Tuberculosis and Lung Disease, EuropeRegion. Eur Respir J 1999; 14: 978–992.

9 National Tuberculosis and Lung Disease Research Institute/World Health Organization Collaborating Centre for Tuberculosis.Report on the Second Meeting of National TB Programme managers from Central and Eastern Europe and the former USSR. Bulletin No 3. Warsaw, Poland: WHO Collaborating Centre for Tuberculosis, 1997: 1–30.

10 American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America. Treatment of tuberculosis. Am J Respir Crit Care Med 2003; 167: 603–662.

11 World Health Organization. Global tuberculosis control. WHO Report 1999. WHO/CDS/CPC/TB/99.259. Geneva, Switzerland: WHO, 1999.

12 Farah M G, Tverdal A, Steen T W, Heldal E, Brantsaeter A B, Bjune G. Treatment outcome of new culture positive pulmonary tuberculosis in Norway. BMC Public Health 2005; 5: 14.735–739.

Table 1, on the next page, presents the characteristics of the 399 patients eligible for this study.1

**Question 1**

What type of study design is described in the abstract? (20 pts)

1. Observational
2. Case Control
3. Retrospective
4. Cross Sectional
5. a and c

e

**Question 2**

What proportion of patients was successfully treated? (20 pts)

68.9

**Question 3**

Calculate a 95% Confidence Interval (CI) for the true population proportion with successful treatment. Hint: The SE of p is the square root of (pq)/n. (40 pts)

SE = sqrt[0.689(1 – 0.689)/399] = sqrt[0.689(0.311)]399 = sqrt[0.214279/399] = 0.0230

UP CI = 0.689 + 1.96 (0.0230) = 0.689 + 0.04508 = 0.73408

LL CI = 0.689 – 1.96 (0.0230) = 0.689 – 0.04508 = 0.64392

Upper limit CI = 0.73408 (20 pts)

Lower Limit CI = 0.64392 (20 pts)

**Question 4**

Which of the following is true with regard to the confidence interval computed in Question 3 above: (25 pts)

1. 95 times out of 100 one would expect a the sample of 399 taken from the same population to have a proportion of successfully treated patients to be between the upper and lower limits of the confidence interval computed in Question 2.
2. 95 times out of 100 one would expect a the sample of 399 taken from the same population to have a proportion of successfully treated patients to be outside the upper and lower limits of the confidence interval computed in Question 2.
3. 5 times out of 100 one would expect a the sample of 399 taken from the same population to have a proportion of successfully treated patients to be outside the upper and lower limits of the confidence interval computed in Question 2.
4. a and c.

**d**

**Question 5** (30 pts; each cell = 5 pts)

Using the information from Table 1, construct a 3 x 2 table to test the association between DOT status and successful treatment.

**Observed Treatment Success**

|  |  |  |  |
| --- | --- | --- | --- |
| DOT | Yes | No | Total |
| Yes | 186 | 64 | 250 |
| No | 86 | 60 | 146 |
| Unknown | 3 | 0 | 3 |
| Total | 275 | 124 | 399 |

Generate the **expected values** for the empty cells below. Hint: the expected value for any cell is the row total x column total divided by the grand (overall) total. (20 pts; each cell = 5 pts)

**Expected Treatment Success Values**

|  |  |  |  |
| --- | --- | --- | --- |
| DOT | Yes | No | Total |
| Yes | 172.3057 | 77.6942 | 250 |
| No | 100.6265 | 45.3734 | 146 |
| Unknown | 2.0676 | 0.9323 | 3 |
| Total | 275 | 124 | 399 |

**Question 6**

Using the DOT status groups, generate the chi-squared test statistic, by hand, using a calculator, or using a computer.

Alpha = 0.05

df = \_\_\_\_\_ (5 pts)

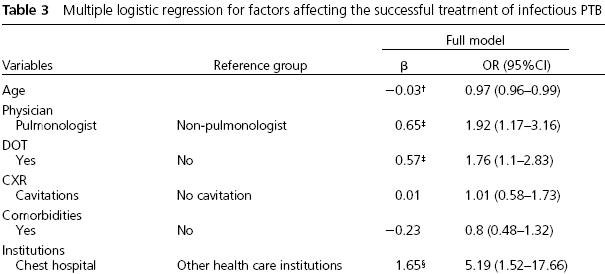
Critical value = \_\_\_\_\_ (5 pts)

Chhi-squared test statistic = \_\_\_\_\_ (5 pts)

Based on comparing the Chi square statistic to the critical value which of the following is true? (5 pts)

1. Successful outcome is dependent on DOT status.
2. Successful outcome is independent of DOT status.
3. No conclusion can be made.
4. The Chi square test is invalid because of only 1 degree of freedom.

Multiple logistic regression analysis allows us to look at the impact of independent variables (potential predictor variables) on a dichotomous outcome variable such as successful treatment completion (yes/no) when controlling for other independent variables. Table 3 presents some of the results of the multiple logistic regression model.1 The outcome is successful treatment.



Note: 

Other footnotes are intentionally excluded from this table.

One way to assess the importance of a potential predictor variable is to examine the odds ratios (ORs) and associated 95% CIs that are estimated from the logistic regression model.

**Question 7**  (25 pts)

Which independent variables listed below is (are) positively significantly associated with successful treatment?

1. Institutions
2. Physician
3. DOT
4. CXR
5. a, b, and c.

**References**

1.Chung,\*† Y-C. Chang,† M-C. Yang†, \* Department of Internal Medicine, Hualien General Hospital, Hualien, † Institute of Health Care Int J Tuberc Lung Dis 11:59–64 © 2007 The Union

2. Dawson, B and Trapp, R **Basic &Clinical Biostatistics**, 4th edition, Lange Basic Science, 2004 page 152.