BIOSTAT Case Study 2: Tests of Association for Categorical Data

Time to Complete Exercise: 45 minutes

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.

ASPH BIOSTATISTICS COMPETENCIES ADDRESSED

A.1. Describe the roles biostatistics serves in the discipline of public health
A. 3. Describe preferred methodological alternatives to commonly used statistical methods when assumptions are not met
A.5. Apply descriptive techniques commonly used to summarize public health data
A.6. Apply common statistical methods for inference
A. 9. Interpret results of statistical analyses found in public health studies

Please evaluate this material by clicking here:
http://www.zoomerang.com/Survey/?p=WEB229TEWRZQ7B

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 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 and 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 INSTRUCTORS' GUIDE Version 1.0.

Statistical output provided in this exercise was generated using JMP 7.0, SAS Institute Inc.

Date Last Modified: November 16, 2009

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]

Question 1
What type of study design is described in the abstract?

Answer Key
This is an historical or retrospective cohort study. The cohort was assembled based on date of entry into the TB registry and then medical records were reviewed for 15 months after PTB was confirmed.

Following is the methods section of this article[1].

> **METHODS**
> We carried out a population-based medical record review in southern Taiwan, where the only chest specialty hospital geared towards specialised 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 centres, 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 centres' 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 χ^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 $P \leq 0.10$ 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.

Statistical output provided in this exercise was generated using JMP 7.0, SAS Institute Inc.

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.

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]

Table 1 Characteristics of 399 patients with PTB and univariate analyses of potential predictors of successful treatment

Characteristics	Patients *n* (%)	Successful treatment* *n* (%)	*P* value†
Sex			0.392
Male	293 (73.4)	198 (67.6)	
Female	106 (26.6)	77 (72.6)	
Patients with comorbidities			0.018
No comorbidity	143 (35.9)	109 (76.2)	
With comorbidities	255 (64.1)	165 (64.7)	
Unknown	1	1 (100)	
CXR			0.015
Cavitations	127 (31.9)	98 (77.2)	
No cavitation	271 (68.1)	176 (64.9)	
Unknown	1	1 (100)	
Previous TB treatment			0.815
Yes	22 (5.8)	16 (72.7)	
No	360 (94.2)	246 (68.3)	
Unknown	17	12 (70.6)	
DOT			0.002
Yes	250 (63.1)	186 (74.4)	
No	146 (36.9)	86 (58.9)	
Unknown	3	3 (100)	
Diagnostic and treating physicians			<0.001
Pulmonologist	298 (74.7)	221 (74.2)	
Non-pulmonologist	102 (25.6)	55 (53.9)	
Health-care institutions			<0.001
Medical centre	105 (26.3)	68 (64.8)	
Regional hospital	144 (36.1)	92 (63.9)	
District hospital and primary practitioners	98 (24.6)	66 (67.3)	
Chest specialty hospital	52 (13.0)	49 (94.2)	
Total	399 (100)	275 (68.9)	

* 15-month follow-up after start of treatment.
† χ^2 test.
PTB = pulmonary tuberculosis; CXR = chest X-ray; TB = tuberculosis; DOT = directly observed treatment.

Question 2
What proportion of patients was successfully treated?

Answer Key
275/399= 68.9%

Question 3
Calculate a 95% Confidence Interval (CI) for the true population proportion with successful treatment. Hint: The SE of p is square root of (pq)/n.

Answer Key
95% CI for $p = \hat{p} \pm 1.96 \times \text{se of } \hat{p}$
SE p = sqrt of (pq)/n = sqrt (0.689 x 0.311)/399) = .0232

Distributions
Treatment Success
CIs

Level	Count	Prob	Lower CI	Upper CI	1-Alpha
No	124	0.31078	0.267351	0.357812	0.950
Yes	275	0.68922	0.642188	0.732649	
Total	399				

Note: Computed using score Confidence Intervals.

Question 4
Write a sentence that describes the meaning of the 95% CI.

Answer Key
We can be 95% confident that the interval ranging from 0.64 to 0.73 covers the true population proportion of patients with successful treatment (allowing for rounding).

Question 5

Using the information from Table 1, construct a 2x2 table to test the association between DOT status and successful treatment.

Observed Treatment Success

DOT	Yes	No	Total
Yes			250
No			146
Unknown			3
Total	275	124	399

Answer Key

Observed Treatment Success

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

Question 6

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.

Expected Values Treatment Success

DOT	Yes	No	Total
Yes			250
No			146
Unknown			3
Total	275	124	399

Answer Key

DOT by Treatment Success

Count **Expected**	**Yes**	**No**	
Yes	**186** **172.306**	**64** **77.6942**	**250**
No	**86** **100.627**	**60** **45.3734**	**146**
Unknown	**3** **2.06767**	**0** **0.93233**	**3**
Total	**275**	**124**	**399**

Question 7

Given these expected values, is the chi-square test an appropriate statistical test?

Answer Key

No (2 cells have expected values of <5).

Question 8

Which of the following would be an appropriate approach to the analysis of these data?

Yes No A. Fisher's Exact Test using all data

Yes No B. Chi-square or Fisher's Exact testing after deleting unknowns

Answer Key **A. Yes B. Yes**

Statistical output provided in this exercise was generated using JMP 7.0, SAS Institute Inc.

Question 9

Using the groups with known DOT status, generate the chi-squared test statistic, by hand, using a calculator, or using a computer. Hint: the short-cut formula for a 2x2 table[2] is:

$$\chi^2 = \frac{n(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

Where a,b,c, d and n are shown in this table:

	Outcome		
Independent variable	Yes	No	Total
Yes	a	b	a+b
No	c	d	c+d
Total	a+c	b+d	a+b+c+d=n

Answer Key

Contingency Table

DOT By Treatment Success

Count Total % Col % Row %	Yes	No	
Yes	186 46.97 68.38 74.40	64 16.16 51.61 25.60	250 63.13
No	86 21.72 31.62 58.90	60 15.15 48.39 41.10	146 36.87
Total	272 68.69	124 31.31	396

Test	Chi-square	Prob>Chi-square
Pearson	10.290	0.0013

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9995	Prob(Treatment Success=no) is greater for DOT=yes than no
Right	0.0011	Prob(Treatment Success=no) is greater for DOT=no than yes
2-Tail	0.0016	Prob(Treatment Success=no) is different across DOT

Question 10

Statistical output provided in this exercise was generated using JMP 7.0, SAS Institute Inc.

Is DOT status related to successful treatment outcome?

Answer Key
Yes. Using a 2-sided test, the probability of treatment success is statistically significantly higher for those on DOT versus those not on DOT.

Question 11
What is the p value associated with the chi-squared test?

Answer Key
$p < 0.05$ by the χ^2 test with one degree of freedom, excluding unknown DOT status.

Question 12
Write a sentence that describes the meaning of the p value.

Answer Key
The difference between the observed proportions that were successful by DOT status (or one even more extreme) is unlikely to have been observed by chance alone. A difference of this magnitude or more extreme would occur less than 5 times out of 100 by chance alone.

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.

Table 3 Multiple logistic regression for factors affecting the successful treatment of infectious PTB

		Full model	
Variables	Reference group	β	OR (95%CI)
Age		−0.03†	0.97 (0.96–0.99)
Physician			
Pulmonologist	Non-pulmonologist	0.65‡	1.92 (1.17–3.16)
DOT			
Yes	No	0.57‡	1.76 (1.1–2.83)
CXR			
Cavitations	No cavitation	0.01	1.01 (0.58–1.73)
Comorbidities			
Yes	No	−0.23	0.8 (0.48–1.32)
Institutions			
Chest hospital	Other health care institutions	1.65§	5.19 (1.52–17.66)

Note:
PTB = pulmonary tuberculosis; OR = odds ratio; CI = confidence interval; DOT = direct observation of treatment; CXR = chest X-ray.
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 13
If there were no association between the independent variables and the outcome in a logistic regression model, what would you expect the estimated β values, ORs, and 95% CIs from the model to look like?

Answer Key
The estimated β values would be close to 0, the estimated ORs would be close to 1, and 95% CIs for the ORs would include the number 1.

Question 14
Which independent variables in the full model in Table 3 are statistically significantly associated at the 0.05 significance level with successful treatment completion of infectious PTB?

Answer Key
Age, being seen by a pulmonologist, receiving DOT, being treated in a chest hospital

Question 15
Which independent variables are positively significantly associated with successful treatment?

Answer Key
Being seen by a pulmonologist, receiving DOT, being treated in a chest hospital

Question 16
Which independent variables are negatively significantly associated with successful treatment?

Answer Key
Age

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.