

Final exam, 8 December 2009

This question for the final exam is worth 6% of the total course mark. It has two parts, labeled *a)* and *b)*, that should both be answered.

The data for this question originate from a study of leprosy in the Karonga District, Northern Malawi (in South-East Africa), carried out in the 1980s. In this area, Bacillus Calmette-Guérin (BCG) vaccination against tuberculosis had been offered for a prolonged time to the entire population as a standard preventive measure. The aim of this particular part of the study was to investigate whether BCG vaccination in early childhood confers any protection against leprosy, which is caused by a closely related bacillus. New cases of leprosy reported during a certain period were examined for presence or absence of the characteristic scar left by BCG vaccination. The researchers were faced with the choice between relying on a health survey of the population during approximately the same period which included examination for the BCG scar, or drawing a sample from the population to be examined specifically for the BCG scar (and leprosy). Although the survey was intended to cover the entire population, some inaccuracies would be inevitable in practice. An additional variable of interest was age, here restricted to only two age categories: less than 25 years old, and age 25-35 years. No persons over the age of 35 years at the time of the study were included, because they would have been children at a time when BCG vaccination was not widely available.

The data in the table below give counts of persons in different categories defined by the variables `age` (0: <25 years, 1: 25–35 years), `scar` (0: absent, 1: present), and `leprosy` (0: absent, 1: new case). The tables includes both the counts for the survey of the non-diseased population, and the counts for the sample drawn from this population. (*Note:* In reality, only one of these two data sources were available, but it is not important for these questions to know which one was actually used.)

| Leprosy category | Age category | | | |
|---------------------------|-----------------|------------|----------------------|------------|
| | 1 (25–35 years) | | 0 (\leq 24 years) | |
| | Scar category | | Scar category | |
| | 1 (present) | 0 (absent) | 1 (present) | 0 (absent) |
| 1 (diseased) | 17 | 83 | 84 | 76 |
| 0 (non-diseased) — survey | 2859 | 9601 | 43169 | 24993 |
| 0 (non-diseased) — sample | 100 | 300 | 426 | 214 |

Part a): Survey data

Assume the researchers decided to use the data obtained from the survey. Describe the study design, and draw a graph to represent the causal structure between the variables involved in the dataset you find most reasonable (prior to any data analysis). Next, carry out an epidemiological analysis to determine the epidemiological roles of the different variables. Use the attached Stata listings for Part *a)*, and supplement with any calculations of measures

of association you find necessary to support your analysis. (*Note:* You are not expected to manually calculate any significance tests, so if you need an assessment of statistical significance, use your judgement based on the information available and explain how you would obtain a precise statistical assessment using statistical software.) As a summary of your analysis, give one (or multiple) measure(s) of association that you think best describe(s) the relation between BCG vaccination and leprosy, and draw conclusions.

Part b): Sample data

Assume the researchers decided to use the data obtained from the sample. Carry out a similar epidemiological analysis as in Part a), answering to the same points as above. If some of your answers are identical to those of a), you may just note this without duplicating the answer. (*Hint:* You should start by characterising the study design and the sampling mechanism.)

Stata listing for Part a)

```
. cs leprosy scar [fw=nsurvey], by(age) or
```

| age | OR | [95% Conf. Interval] | | M-H Weight |
|--------------|----------|----------------------|----------|----------------------|
| 0 | .6399 | .4695909 | .8719752 | 48.02032 (Cornfield) |
| 1 | .6878174 | .4098706 | 1.154303 | 18.89307 (Cornfield) |
| -----+ | | | | |
| Crude | .4774225 | .3721376 | .612495 | |
| M-H combined | .6534295 | .4995528 | .8547047 | |

```
-----+-----
Test of homogeneity (M-H)      chi2(1) =    0.054  Pr>chi2 = 0.8157

      Test that combined OR = 1:
                Mantel-Haenszel chi2(1) =    9.85
                Pr>chi2 =    0.0017

. cs leprosy age [fw=nsurvey], by(scar) or
```

| scar | OR | [95% Conf. Interval] | | M-H Weight |
|--------------|----------|----------------------|----------|----------------------|
| 0 | 2.842932 | 2.083411 | 3.879337 | 20.99606 (Cornfield) |
| 1 | 3.055818 | 1.82243 | 5.12414 | 5.206183 (Cornfield) |
| -----+ | | | | |
| Crude | 3.419041 | 2.662925 | 4.389855 | |
| M-H combined | 2.885231 | 2.200974 | 3.782214 | |

```
-----+-----
Test of homogeneity (M-H)      chi2(1) =    0.055  Pr>chi2 = 0.8146

      Test that combined OR = 1:
                Mantel-Haenszel chi2(1) =   64.71
                Pr>chi2 =    0.0000
```

Stata listing for part b):

```
. cs leprosy scar [fw=nsample], by(age) or
```

| age | OR | [95% Conf. Interval] | | M-H Weight |
|--------------|----------|----------------------|----------|-------------------|
| 0 | .5552261 | .3911851 | .7880252 | 40.47 (Cornfield) |
| 1 | .6144578 | .3500458 | 1.07949 | 16.6 (Cornfield) |
| -----+ | | | | |
| Crude | .6207284 | .4705759 | .8188119 | |
| M-H combined | .5724549 | .4240836 | .7727357 | |

```
Test of homogeneity (M-H)      chi2(1) =    0.089  Pr>chi2 = 0.7659
```

```
Test that combined OR = 1:
```

```
      Mantel-Haenszel chi2(1) =    13.54  
      Pr>chi2 =    0.0002
```

```
. cs leprosy age [fw=nsample], by(scar) or
```

| scar | OR | [95% Conf. Interval] | | M-H Weight |
|--------------|----------|----------------------|----------|----------------------|
| 0 | .7790351 | .5455068 | 1.112504 | 33.87816 (Cornfield) |
| 1 | .8621429 | .4930408 | 1.508729 | 13.39713 (Cornfield) |
| -----+ | | | | |
| Crude | 1 | .7566705 | 1.321621 | |
| M-H combined | .8025866 | .5929594 | 1.086323 | |

```
Test of homogeneity (M-H)      chi2(1) =    0.088  Pr>chi2 = 0.7662
```

```
Test that combined OR = 1:
```

```
      Mantel-Haenszel chi2(1) =     2.04  
      Pr>chi2 =    0.1532
```