Methods in Epidemiologic Research

Sample Problems

Chapter 6 – Measures of Association

Preparation

We will carry on with the data that were used in the exercises for Chapter 4, but we will focus on the effect of having had a previous myocardial infarction (-prmi-) on the risk of dying within 1 year of admission for the MI included in this dataset. All individuals in this dataset were followed for at least 1 year after admission, so there are no withdrawals (censored observations) to worry about.

The first thing you will have to do is generate three new variables.

- Create a 0/1 variable which indicates whether or not the individual died during the 1st year following admission.

If you are using Stata, some code to generate the variable is

```
gen obs_1yr=(surv_mi+1)/365.25
replace obs_1yr=1 if obs_1yr>1
```

- Create a variable indicating the amount of time that the person was under observation. This will be 1 year if the person survived and for those who died, it will be the time until death. When computing this variable, take the following 2 issues into consideration.
  - Before computing this variable, add 1 to the original survival time variable (-surv_mi-) because people who died on the day of admission were listed as having a "survival" of 0 days, but be want them to have a "time under observation" of 1 day.
  - Convert the time from days to years (by dividing by 365.25). This is just to avoid having very small values when estimating rates.

If you are using Stata, some code to generate the variable is

```
gen died_1yr=0
replace died_1yr=1 if died==1 & obs_1yr<1
```

- Generate a 0/1 variable indicating whether or not the person survived the full 5 year period. (This is just done to confirm that there were no withdrawals during the year.)

```
gen obs_full1yr=obs_1yr==1
tab obs_full1yr died_1yr
```

Questions

1. Compute and interpret the risk ratio (RR) for previous myocardial infarction.

```
. cs died_1yr prmi, exact /* P-value is Fisher's exact P */

                 | previous MI            |
                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+-----------
           Cases |       287         490  |        777
        Noncases |       548        1640  |       2188
-----------------+------------------------+-----------
```

```
           Total |       835        2130  |       2965
                 |                        |
            Risk |  .3437126    .2300469  |   .2620573
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
 Risk difference |         .1136656       |    .0768253     .150506
      Risk ratio |         1.494098       |    1.322842    1.687523
 Attr. frac. ex. |         .3306996       |     .244052    .4074156
 Attr. frac. pop |         .1221503       |
                 +-----------------------------------------------
                                  1-sided Fisher's exact P = 0.0000
                                  2-sided Fisher's exact P = 0.0000
```

The RR was 1.49 meaning that individuals who had had a previous MI were about 50% more likely to die during the 1st year after this admission than an individual who had not had a previous MI. The P-value (2-sided) is highly significant, suggesting that the apparent association was not likely due to chance.

2. Compute and interpret the incidence rate ratio (IR) for previous myocardial infarction.

```
. ir died_1yr prmi obs_1yr
```

```
                 | previous MI            |
                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+------------
        died_1yr |       287         490  |        777
         obs_1yr |  615.8138    1744.438  |   2360.252
-----------------+------------------------+------------
                 |                        |
  Incidence rate |  .4660499    .2808927  |   .3292022
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
 Inc. rate diff. |         .1851572       |    .1257789    .2445355
 Inc. rate ratio |         1.659174       |    1.429204    1.923314 (exact)
 Attr. frac. ex. |         .3972905       |    .3003098    .4800641 (exact)
 Attr. frac. pop |         .1467469       |
                 +-----------------------------------------------
                     (midp)    Pr(k>=287) =                  0.0000 (exact)
                     (midp) 2*Pr(k>=287) =                  0.0000 (exact)
```

The IR was 1.66 meaning that the rate at which deaths were occurring in individuals who had had a previous MI was about 66% higher than the rate in individuals who had not had a previous MI. As with the RR, the P-value (2-sided) is highly significant, suggesting that the apparent association was not likely due to chance.

3. Compute and interpret the odds ratio (OR) for previous myocardial infarction.

```
. cc died_1yr prmi, exact
```

```
                                                        Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+------------------------
           Cases |       287         490  |        777       0.3694
        Controls |       548        1640  |       2188       0.2505
-----------------+------------------------+------------------------
           Total |       835        2130  |       2965       0.2816
                 |                        |
```

```
                    |      Point estimate    |    [95% Conf. Interval]
                    |------------------------+------------------------
         Odds ratio |          1.752868      |    1.465919    2.095881 (exact)
    Attr. frac. ex. |          .4295062      |    .3178339    .5228736 (exact)
    Attr. frac. pop |          .1586465      |
                    +-----------------------------------------------
                                    1-sided Fisher's exact P = 0.0000
                                    2-sided Fisher's exact P = 0.0000
```

The OR was 1.75 meaning that individuals who had had a previous MI had an odds of dying during the 1st year after this admission that was 75% higher than the odds for an individual who had not had a previous MI. OR are often loosely interpreted as RR, but in this case you can see that it is not a good estimate of the RR (see part (b) below). As with the RR, the P-value (2-sided) is highly significant, suggesting that the apparent association was not likely due to chance.

(a) Does the relative magnitude of the RR, IR and OR make sense?

Yes. The usual order is OR>IR>RR. The relationship between OR and RR is discussed below in part (b). The relationship between IR and RR is not as predictable as it also depends on how much follow-up time was observed in the exposed and unexposed groups.

(b) Is the OR close to the RR? Why, or why not?

No they are not particularly close (1.75 vs 1.5) because the outcome is not rare. 26% of the individuals (77/2965=0.26) died during the year. As the frequency of the outcome goes up, so to does the difference between the OR and the RR.

4. Compute and interpret the following measures of effect.

All of the calculations are provided in the output above.

(a) Risk difference (RD)

The RD was 0.11 meaning that among 100 people who had had a previous MI, there would be 11 additional deaths compared with what there would have been if those people had not had a previous MI (34 deaths compared to 23).

(b) Incidence rate difference (ID)

The incidence rate difference (ID) was 0.19 meaning that for 100 person-years of follow-up time, we would observe 19 more deaths if the individuals followed had had a previous MI compared with a comparable group who had not had a previous MI (47 deaths per 100 person-years vs 28).

(c) Attributable fraction – exposed (Af_e)

The Af_e was 0.33 meaning that among exposed individuals, 33% of deaths could be attributed to the fact those individuals had been exposed. To relate this back to the risks in the two groups … we had 11 additional deaths (see RD in part (a)) and these 11 deaths account for 33% of the deaths in the exposed group (11/34=0.32). (Note – it is also possible to compute the Af_e based on the incidence rate or odds data and these values are large (0.40 and 0.43 respectively), but it is most common to base Af_e on risk data.)

(d) Population attributable risk (PAR)

The PAR is the only parameter which is not computed directly by Stata. It is calculated as the risk in the whole population minus the risk in the non-exposed group (0.26 – 0.23 = 0.03). This means that

among 100 people in this population, 3 of the deaths occurring within 1 year of admission could be attributed to the fact that some people had had previous MI.

(e) Population attributable fraction (Af_p)

The Af_p was 0.12 meaning that in this population, 12% of deaths could be attributed to the fact some individuals in the population had been exposed (ie had had a previous MI). To relate this back to the risk estimates … we had 3 additional deaths attributable to previous MI (see PAR in part (d)) and these 3 deaths account for 12% of the deaths in the population (3/26=0.12). (Note – it is also possible to compute the Af_p based on the incidence rate or odds data and these values are larger (0.15 and 0.16 respectively), but it is most common to base Af_p on risk data.)

5. Interpret the confidence intervals (CI) for each of the above.

We will just focus on the CI for the RR and leave it to you to look at the other CI. The RR is 1.49 and the CI is from 1.32 to 1.69 meaning that our "best guess" as to the true value of the RR is 1.49, but values between 1.32 and 1.69 would be reasonable estimates. More specifically, if we repeated this study many times in the same population (how boring would that be?) the range of values with the CI would capture the "true value" 95% of the time and the true value would be outside that range 5% of the time.

One commonly hears that the 95% CI means that there is a 95% probability that the true value lies within the CI. This is not actually correct. For any single study, the true value is either within the CI (probability =100%) or it is not (probability=0%).

You will also have noted that there are no estimates of the CI for the Af_p. There are no readily available formulae for these calculations.