

Course EPIB-681: Data Analysis II [Winter 2004]

Assignment 3

material in www.epi.mcgill.ca/hanley/c607/ unless otherwise specified
(username: c607 ; password: 8 letters, H**J*## both case-sensitive)material in

• From "Questions" on pp166-167 at end of Chapter 8 of Rothman 2002

- 1 (crude risk ratio)
- 2 (pooled risk ratio)
- 4 (SMR)
- 8 (contribution from a stratum in which all subjects were unexposed)

• "Homegrown"

- 1 In his question 8, Rothman asks you to *imagine* a stratum. Which strata in the *real* example used in the 1959 classic article by Mantel and Haenszel are of this type? The relevant table is on p 738 (in 'part 2' of the M-H article -- under 'Resources for Chapter 9' in the course 607 website).
- 2 What are the contributions of a stratum to the summary odds ratio if (i) all of the 'controls' in the stratum are unexposed and all of the cases in the stratum are exposed (ii) the converse? Which strata in the Table 1 are of these types?
- 3 Mantel and Haenszel do not refer to the 1955 paper by Barnet Woolf (also under Resources for Chapter 9 in the course 607 website). How well does Woolf's method work if strata have sparse data (see the frequencies used in the examples in Woolf's paper)?

Compute a M-H summary odds ratio, a test-based* 95%CI, and a Robins-Greenland-Breslow CI, from the 4 age-strata of "housewives" in Table 1 of the M-H article [Mantel and Haenszel have already done most of the work for you for the M-H summary OR measure, and for the test-based CI, even though that method hadn't yet been developed in 1959; the 'Robins-Greenland-Breslow' formula (the last variance formula in Rothman's Table 8-4) for the variance of the log of the M-H summary OR estimate was developed later still -- in 1986.

{ * as is common practice nowadays, omit the continuity correction in the X^2 statistic; this avoids the paradoxical situation where even if A equaled $E[A]$ in *each* stratum -- so that each (stratum-specific) OR estimate (and thus the summary OR estimate would be unity -- the X^2 value would be nonzero! }

Compute the Woolf summary odds ratio and associated 95% CI from the same 4 'housewives' strata.

Comment on (i) the relative complexity of the calculations (you should do them -- this once at least -- by hand or by calculator, or program them in a spreadsheet) and on (ii) which method is more versatile (i.e. handles strata of all sizes.. from ones like in the Woolf examples, down to strata consisting of matched pairs... Mantel and Haenszel discuss this issue on page 741)

Test-based CIs are accurate near the *null* value (since the standard error is computed under the null hypothesis) and are less so for parameter values away from the null. Mantel and Haenszel's example of lung cancer and smoking provides an opportunity to see how much they differ in practice in an extreme non-null situation, and whether the 'new and improved' Robins-Greenland-Breslow CI is very different. Compare the two in this example. Is Stata's strong warning "We recommend that test-based confidence intervals be used only for pedagogical purposes and never be used for research" warranted in this example?

{later we will see a *conditional* approach to finely stratified data, using conditional logistic regression}

4 Refer to the article A POPULATION-BASED STUDY OF MEASLES, MUMPS, AND RUBELLA VACCINATION AND AUTISM [under **Nov. 11 lecture in med2**, reachable from link at top of 626 website: username med2, password: same as for the cxxx epidemiology courses] and to the righthand column of page 3 of the med2 handout of Nov. 11, dealing with the reason for the big difference between the *crude* rate ratio of 1.44 and the '*adjusted*' rate ratio of 0.92.

- Explain to a *journalist* why the big difference between these estimates in this example.

[The link **How could one get a crude Rate Ratio (RR) of 1.45?** gives a bit more detail]

- Complete the exercise at the bottom of the second page in the link **Why can the crude Rate Ratio (RR) be 1.45 if RR=1 at all ages and in all years?** Hint: use each vertical slice as a 'stratum' and use the Mantel-Haenszel summary rate ratio (Sum over 8 vertical slices*) and calculate a (test-based) CI to accompany it.

for one stratum...

$$\text{rate ratio} = \frac{\# \text{Exposed Cases} / \text{Exposed P-T}}{\# \text{Unexposed Cases} / \text{Unexposed P-T}} = \frac{\# \text{Exposed Cases} \times \text{Unexposed P-T}}{\# \text{Unexposed Cases} \times \text{Exposed P-T}}$$

aggregating (summing) over 2 or more strata... (see Rothman2002 p 153)

$$\text{MH summary rate ratio} = \frac{\text{Sum} [\# \text{Exposed Cases} \times \text{Unexposed P-T} / \text{Combined P-T}]}{\text{Sum} [\# \text{Unexposed Cases} \times \text{Exposed P-T} / \text{Combined P-T}]}$$

-
- * *Ideally, should summarize over all (36!) age-year cells, to also keep calendar year 'constant'. Since focus here is on how the calculations work, ignore (collapse over) calendar time for now.*

To those doing calculations manually: spreadsheets were made for repetitive calculations like these! [if you want a quickstart, the spreadsheet under "Other resources" in Chapter 9 of course 607 is a useful template to build on]

Or, you can modify the SAS or the Stata example in the Resources for Chapter 9.

To save time and to standardize on one observer/counter, JH has manually counted the numbers of cases, and calculated the person-time, for each of the age-calendar-year 'cells' shown on the diagram. These, along with statements that will do several analyses, on the 626 web page, can be found in the file "counts / person times measured from diagram, together with SAS program" under "Regression for Rates and proportions" in the 626 course page.

The task in the 626 assignment that this is taken from was to run each analysis (they are labeled using title statements) and interpret and contrast the outputs from the different regression analyses. We will come to these regression analyses later. For now, you might just want to run the Mantel-Haenszel analysis in SAS or Stata, or (manually or by software) add up the numbers of child-years and the numbers of events in the same age-slice to collapse the finer calendar time-age strata into 'age-only' strata.