STAT 6200 — Design of Experiments for Research Workers
Lab 2 – Due Tuesday, September 21

Example — Oral Contraceptive Use and Heart Attack:

In class, we gave an example of a fictional case-control study to investigate the relationship between oral contraceptive (OC) use and heart attack (myocardial infarction or MI) among women. Suppose instead that the study had been conducted as a prospective cohort study, in which 5000 OC users and 10000 non-OC users were identified and then followed for 3 years to observe the heart attack incidence in the two groups. The results were as follows:

		Heart Attack: Yes	Heart Attack: No	
OC Use	Yes	13	4987	5000
	No	7	9993	10000
		20	14980	15000

These data are contained in the minitab worksheet OCMI.MTW. Please download this file from the course web site, and open it in minitab. In this worksheet, the data are grouped. That is, rather than listing OC use and MI status for each of 15000 subjects (15000 rows in the data set), we have included rows only for the four combinations of OC use and MI status plus a frequency variable summarizing how many women were observed at each combination of these variables.

To generate the contingency table given above for this example, select

Stat > Tables > Cross Tabulation and Chi-Square

Now select variable OC Use under "For rows:", select variable Heart Attack under "For columns:", and select variable Frequency under "Frequencies are in:". Place a check next to "Counts", "Row percents", "Column percents", and "Total percents" and then click OK.

Minitab should now give a contingency table like that given above for these data. Notice that Minitab orders the rows and columns alphabetically, so the order of the rows above are switched. Also notice that Minitab gives row percents, column percents, and total percents.

In this study design the row totals are fixed by design because a specific number of OC users and non-OC users were selected for study. This means that we can estimate the conditional probability of a MI given that the woman is an OC user and the conditional probability of a MI given that the woman is not an OC user from the row percents. That

is, the risk of MI for OC users (the exposed group) is estimated to be

$$\widehat{\text{Risk for exposed}} = \frac{13}{5000} = 0.0026$$

or .26%, and the risk of MI for non OC users (the unexposed group) is estimated to be

$$\widehat{\text{Risk for unexposed}} = \frac{7}{10000} = 0.0007$$

or 0.07%.

Therefore, the risk difference is estimated to be

$$\widehat{\text{risk difference}} = \widehat{\text{Risk for exposed}} - \widehat{\text{Risk for unexposed}} = 0.0026 - 0.0007 = 0.0019,$$

or 0.19%, and the risk ratio is estimated to be

$$\widehat{RR} = \frac{\widehat{\text{Risk for exposed}}}{\widehat{\text{Risk for unexposed}}} = \frac{0.0026}{0.0007} = 3.71.$$

The odds of MI in the OC group are

$$\text{odds for exposed} = \frac{\text{Risk for exposed}}{(1 - \text{Risk for exposed})}$$

which is estimated to be

$$\widehat{\text{odds for exposed}} = \frac{.0026}{1 - .0026} = .00260678$$

and the odds of MI in the non-OC group are

$$\text{odds for exposed} = \frac{\text{Risk for exposed}}{(1 - \text{Risk for exposed})}$$

which is estimated to be

$$\widehat{\text{odds for exposed}} = \frac{.0007}{1 - .0007} = .00070049$$

Therefore, the odds ratio is estimated to be

$$\hat{OR} = \frac{.00260678}{.00070049} = 3.72$$

Notes from this example:

- Notice that the estimated OR for a 2×2 table like this one can always be computed as the ratio of the products of the cell counts on each diagonal of the table. E.g., in this example,

$$\hat{OR} = \frac{(13)(9993)}{(7)(4987)} = 3.72$$

  For this reason, the OR is sometimes called the cross-product ratio.

- Notice that in this example, the row totals were fixed, allowing us to estimate the conditional probability of disease given exposure status (the risk of disease in each group). This allowed us to estimate the risk difference, RR, and OR. Recall from lecture that in the case-control design we could not estimate risk because the column totals were fixed. However, we could estimate the OR.

  - Thus, the OR can always be estimated, which is one advantage it enjoys over the RR and risk difference.

- Note also that when the disease is rare, the OR is a pretty good approximation to the RR. The reason for this is that when the risk is small (near 0),

$$\text{odds of disease} = \frac{\text{risk}}{1 - \underbrace{\text{risk}}_{\approx 0}} \approx \frac{\text{risk}}{1} = \text{risk}.$$

  We saw that in this example where the risk of heart attack (in the three year window used here) was very low. Recall that we computed the odds in the exposed group to be $.0026/(1-.0026) = .00260678 \approx .0026$ =risk in the exposed group and odds in the unexposed group was $.0007/(1-.0007) = .00070049 \approx .0007$ =risk in the unexposed group. Therefore, in this example and others in which the risk of disease is low, the relative risk can be approximated by the OR even when it cannot be estimated directly:

$$\hat{RR} \approx \hat{OR}$$

Exercise:

A case-control study was conducted involving women who have given birth at least once to investigate the relationship between breast cancer and age at first birth. The records to 3220 breast cancer sufferers (cases) and 10245 controls were examined to determine age at first birth and whether it occurred after age 30 (the exposure). The data from this study are as follows:

		Breast Cancer		
		Case	Control	
Age at 1st birth	≥ 30	683	1498	2181
	< 30	2537	8747	11284
		3220	10245	13465

You may use Minitab or not as you prefer to answer the following questions:

1. Can the risks of breast cancer in the exposed an unexposed groups be estimated here? If so, estimate them.

2. Estimate the odds ratio describing the association between age at first birth and breast cancer.

3. Interpret the odds ratio.

4. Assuming that the overall risk of breast cancer is small, estimate the risk ratio comparing risk of breast cancer between women with first birth after age 30 and women with first birth before age 30. (See the fourth "Note" on the previous page.)