Biost 536: Categorical Data Analysis in Epidemiology
Emerson, Fall 2013

Homework #2
October 10, 2013

1. Suppose we are interested in measuring any association between estrogen use at any time prior to study enrollment (*estrogen==1)* and CVD death within 4 years using the **risk difference (RD)**.

   a. Provide complete statistical inference regarding such an association. (Include point estimates, confidence intervals, and a p value, along with a full interpretation of those quantities.)

      The following estimates were obtained using linear regression:

	Risk Difference (estrogen vs. no estrogen)	95% confidence interval		P > t
Unadjusted	-.025	-.044	-.006	0.01
Stratum 1: No previous disease	-.012	-.032	.008	0.26
Stratum 2: Previous Disease	-0.065	-.15	.02	0.11
Adjusted for Previous disease	-.016	-.035	.003	0.09
Adjusted for previous disease and age	-.010	-.029	.009	0.32

      The absolute risk of dying of CVD within 4 years of recruitment was .025 lower for those who used estrogen than for those who did not (95% CI: -.044, -.006). This risk difference was statistically significant (p = 0.01).

   b. Is there evidence in the dataset that any such effect is modified by a history of prior CVD (as measured by variable *prevdis)*? Provide results of a statistical analysis in support of your answer.

      From the above we know that there is a significant association between CVD death and estrogen exposure. In order to determine whether the effect is modified by previous disease, we would examine this association in the two strata (those with and without previous disease).

      The presence or absence of previous cardiovascular disease does not appear to modify the relationship between estrogen exposure and cardiovascular death. According to the saturated model, among those with

no previous cardiovascular disease, the risk of dying of CVD within 4 years of recruitment was 1.2% (RD .012 95% CI: -.032, .008) lower for those who had used estrogen than for those who had not (95% CI: -3.2%, .01%). This risk difference was not statistically significant (p=0.27). Among those who did have a history of previous cardiovascular disease, the risk of dying was 6.6% lower (RD = 0.06, 95% CI: -.15, .02) for those who took estrogen than those who never did. These two risk differences are not clinically significant and I would not consider this effect modification.

c. Suppose we just want to ignore any such effect modification. Is there evidence in the dataset that any estrogen-CVD mortality association is confounded by a history of prior CVD? Provide results of a statistical analysis in support of your answer.

In order to confound the relationship between estrogen exposure and CVD mortality, the presence of prior CVD would have to be associated with both estrogen exposure and with CVD mortality while not lying on the causal pathway between the two.

Chi square analysis reveals that there is indeed a significant association between previous disease and estrogen exposure ($X^2=26.2$, $p < 0.001$) as well as between previous disease and CVD mortality ($X^2=91.0$, $p < 0.001$), and common sense would lead us to believe that previous disease does not lie on the causal pathway between the two.

The risk difference of CVD death between those who did and did not use estrogen in the unadjusted model above was -.025 (95% CI: -.044, -.006, p=.01), whereas the risk difference of CVD death between those who did and did not use estrogen given they had the same history of CV disease was -.016 (95% CI: -.035, .003., 0.09). Given that previous disease fits the criteria of a a confounder above and the adjustment is pulling the estimate toward the null, I would say that that it is confounding the estrogen-CVD mortality association.

d. Provide complete statistical inference regarding an association between estrogen and CVD mortality after adjustment for a prior history of CVD.

Though there appeared to be a significant risk difference for CVD mortality for those who took estrogen, this risk difference was no longer significant after adjustment for the presence or absence of previous cardiovascular disease. The the risk difference of CVD death between those who did and did not use estrogen given they had the same history of CV disease was small and nonsignficiant at -.016 (95% CI: -.035, .003., p=0.09).

e. Is there evidence in the dataset that the prior disease adjusted analysis of an association between estrogen-CVD mortality is further confounded by age? Provide results of a statistical analysis in support of your answer.

If the association between estrogen and CVD mortality were to be further confounded by age, I would expect there to be an association between age and estrogen exposure as well as age and CVD mortality, given

presence/absence of previous disease. And I do not believe age lies on the causal pathway between the two.

Testing for a difference in mean age of the exposed to the unexposed in those without previous disease reveals that there are significant differences in mean age by estrogen exposure ($t=6.2$, $p<0.001$), and the same is seen when comparing mean age of the exposed to the unexposed in those without previous disease ($t=2.1$, $p=0.04$).

Examining the difference mean age of those who did and did not died of CVD shows that while there are significant differences in mean age by mortality in those with previous disease ($t=-7.2$, $p<0.001$), this difference was not significant in those without previous disease ($t=-1.89$, $p=0.06$).

After controlling for age in the linear regression model, it does appear that age further confounds the association, as further adjusting for age reduces the point estimate for risk difference from -0.016 to -.009, further driving it toward the null.

f. Provide complete statistical inference regarding an association between estrogen and CVD mortality after adjustment for age and any prior history of CVD.

Though there appeared to be a significant risk difference for CVD mortality for those who took estrogen, this risk difference was no longer significant after adjustment for age and the presence or absence of previous cardiovascular disease (RD= -.009, 95% CI -.03, 0.01). This is a negligible difference in absolute risk. Therefore the data do not show a significant difference in absolute risk of death between those who did and did not use estrogen after adjusting for age and previous disease.

2. Answer all parts of problem 1 using the **odds ratio (OR)** as the measure of association.

The following estimates were obtained using logistic regression

	Odds Ratio (estrogen vs. no estrogen)	95% confidence interval		P > t
Unadjusted	0.25	0.08	0.78	0.02
Stratum 1: No previous disease	0.35	0.08	1.43	0.14
Stratum 2: Previous Disease	0.31	0.41	2.34	0.93
Adjusted for Previous disease	0.34	0.11	1.07	0.06
Adjusted for previous disease and age	0.42	0.13	1.35	0.15

a. The odds of estrogen exposure were significantly higher for those who had died versus those who had not died from CVD (OR = 0.25, 95% CI: 0.07, 0.78, p = 0.02)

b. Is there evidence in the dataset that any such effect is modified by a history of prior CVD (as measured by variable *prevdis)*? Provide results of a statistical analysis in support of your answer.

From the above we know that there is a significant association between CVD death and estrogen exposure. In order to determine whether the effect is modified by previous disease, we would examine this association in the two strata (those with and without previous disease).

The presence or absence of previous cardiovascular disease does not appear to modify the relationship between estrogen exposure and cardiovascular death as measured by the odds ratio. The point estimate of the odds ratio for those with previous disease was 0.31 whereas it was 0.35 for those without. This is not a substantial difference and as with the risk difference I do not think effect modification is seen here, either.

c. Suppose we just want to ignore any such effect modification. Is there evidence in the dataset that any estrogen-CVD mortality association is confounded by a history of prior CVD? Provide results of a statistical analysis in support of your answer.

I established in part 1a. that previous disease meets the criteria for a classical confounder.

The odds ratio of estrogen exposure given CVD death in the entire sample was 0.25 (95% CI: 0.07, 0.78, p = 0.02), whereas it closer to 1 and nonsignificant at 0.34 (95% CI: 0.11, 1.07, p=0.06) for those of the same previous disease status. Once again, I believe that confounding is seen when we measure using the odds as our outcome.

d. Provide complete statistical inference regarding an association between estrogen and CVD mortality after adjustment for a prior history of CVD.

Though there appeared to be a difference in the odds of estrogen exposure given CVD mortality, this odds ratio was no longer significant after adjustment for the presence or absence of previous cardiovascular disease. The odds of estrogen exposure given CVD mortality given identical history of CV disease was nonsignficiant at 0.34 (95% CI: 0.11, 1.07, p=0.06).

e. Is there evidence in the dataset that the prior disease adjusted analysis of an association between estrogen-CVD mortality is further confounded by age? Provide results of a statistical analysis in support of your answer.

I showed in part 1d. above that age meets the criteria for a classical confounder between the estrogen-CVD relationship given previous disease status.

After controlling for age in the logistic regression model, it does appear that age further confounds the association, as further adjusting for age substantially changes the point estimate for odds ratio from 0.34 to 0.42, further driving it toward the null.

f. Provide complete statistical inference regarding an association between estrogen and CVD mortality after adjustment for age and any prior history of CVD.

Though there appeared to be a significant reduction in the odds of estrogen exposure given CVD mortality, this odds ratio was no longer significant after adjustment for age and the presence or absence of previous cardiovascular disease (OR 0.42, 95% CI 0.13, 1.35, p=0.15).

3. Answer all parts of problem 1 using the **risk ratio (RR)** as the measure of association. (Note that the Stata `glm` command can be used to effect such analyses.)

The following estimates were obtained using a generalized linear model with family=binomial and a log link function:

	Risk Ratio (estrogen vs. no estrogen)	95% confidence interval		P > t
Unadjusted	0.26	0.082	0.792	0.018
Stratum 1: No previous disease	0.35	0.11	1.08	0.07
Stratum 2: Previous Disease	0.34	0.05	2.35	0.27
Adjusted for Previous disease	0.35	0.11	1.09	0.07
Adjusted for previous disease and age	0.43	0.14	1.34	0.15

a. The risk of CVD death was significant lower for those who had estrogen exposure versus those who had not (RR=0.26, 95% CI: 0.08, 0.79, p = 0.02).

b. Is there evidence in the dataset that any such effect is modified by a history of prior CVD (as measured by variable *prevdis)*? Provide results of a statistical analysis in support of your answer.

From the above we know that there is a significant association between CVD death and estrogen exposure. In order to determine whether the effect is modified by previous disease, we would examine this association in the two strata (those with and without previous disease).

The presence or absence of previous cardiovascular disease does not appear to modify the relationship between estrogen exposure and cardiovascular death as measured by the relative risk. The point estimate of the relative risk for those with previous disease was 0.34 whereas it was 0.35 for those without. This is not a substantial difference and as with the risk difference and odds ratio I do not think effect modification is seen here.

c. Suppose we just want to ignore any such effect modification. Is there evidence in the dataset that any estrogen-CVD mortality association is confounded by a history of prior CVD? Provide results of a statistical analysis in support of your answer.

I established in part 1a. that previous disease meets the criteria for a classical confounder.

The relative risk of estrogen exposure given CVD death in the entire sample was 0.26, 95% CI: 0.08, 0.79, p = 0.02, whereas it closer to 1 and nonsignificant at 0.35 (95% CI: 0.11, 1.09, p=0.07) for those of the same previous disease status. Once again, I believe that confounding is seen when we measure using the relative risk as our outcome.

d. Provide complete statistical inference regarding an association between estrogen and CVD mortality after adjustment for a prior history of CVD.

Though there appeared to be a difference in the relative risk of CVD mortality given estrogen exposure relative risk was no longer significant after adjustment for the presence or absence of previous cardiovascular disease at 0.35 (95% CI: 0.11, 1.09, p=0.07).

e. Is there evidence in the dataset that the prior disease adjusted analysis of an association between estrogen-CVD mortality is further confounded by age? Provide results of a statistical analysis in support of your answer.

I showed in part 1d. above that age meets the criteria for a classical confounder between the estrogen-CVD relationship given previous disease status.

After controlling for age in the generalized linear model, it does appear that age further confounds the association, as further adjusting for age substantially changes the point estimate for odds ratio from 0.34 to 0.43, further driving it toward the null.

f. Provide complete statistical inference regarding an association between estrogen and CVD mortality after adjustment for age and any prior history of CVD.

Though there appeared to be a significant reduction the relative risk of CVD death given estrogen exposure, this relative risk was no longer significant after adjustment for age and the presence or absence of previous cardiovascular disease (RR 0.43, 95% CI 0.14, 1.34, p=0.15).

4. Of the three measures of association used above, how similar were the conclusions? What are the relative advantages and disadvantages of the three?

   Overall, the conclusions were similar whether measuring outcome using RD, RR, or OR. The take home conclusion was that in the unadjusted analyses of all three, estrogen appeared to be protective against CVD mortality, however this effect disappeared when we controlled for previous disease and age. Previous disease did not seem to be an effect modifier in any of the three measures, though the difference in effect was most pronounced when measured with the risk difference.

   This is a cohort study design and the main scientific question we are asking here is whether estrogen use increased the risk of CVD mortality within four years of enrollment. Therefore I think that the relative risk has the advantage of giving a more natural interpretation that is easier to apply clinically.