Erik Parner
22 October 2015

Solution to Exercise 7-2

Background

The standard epidemiological advice of adjusting for a risk factor is: only when the risk factor is associated with the exposure and not on the causal pathway between the exposure and the outcome. This is the standard definition on confounder. This exercise show that the advice should depend on the type of association measure (regression model).

Statistical methods

The mean age and change in Tatsoib between the active drug and placebo group were compared using the t-test. Normality was accessed using QQ-plots and variance homogeneity using the F-test. The proportion of individuals with an increase in Tatsoib were evaluated using the binomial model, and proportions where compared between groups using the χ^2-test. The mean change in Tatsoib between the active drug and placebo group were also compared in an analysis of covariance, adjusting for the age of the individual. The odds of individuals with an increase in Tatsoib were similarly compared in a logistic regression analysis, adjusting for the age of the individual.

Results

1. Make a statistical analysis that confirms this result.

The mean age in the active drug group was 41.0 (96% CI: 37.8;44.2) and in the placebo group 43.5 (CI: 39.5;47.5). There was no statistical significant difference in the means between the two groups (p=0.36). Since the study is a randomized study the analysis really examines that the randomization process, and such hypothesis that the randomization was carried out properly are often of less interest.

2. Describe the change in Tatsoib in each of the two groups.

3. Estimate the effect of the new drug adjusted for a possible placebo effect.
Write a conclusion on the possible effect of the new drug.

The mean change in Tatsoib in the active drug group was 1.13 (CI: 0.46;1.80) and in the placebo group 0.37 (CI: -0.37;1.11). There is thus a 0.76 (CI: -0.22; 1.74) mean higher change in the active drug group as compared to the placebo group, which is however not statistical significant (p=0.13).

4. Perform a relevant statistical analysis of these data.

5. Comment on the similarities and differences between the analyses and results in Questions 3 and 4.

The proportion with an increase in Tatsoib in the active drug group was 74.3% (CI: 56.7;87.5%) and in the placebo group 48.6% (CI: 31.4;66.0%). The was a statistical significant higher proportion with an increase in Tatsoib in the active drug group as compared to the placebo group, a difference of 25.7% (CI:3.7;47.7%, p=0.027), corresponding to an odds ratio of 3.1 (CI: 1.1;8.4).

The analysis in Question 3 and 4 consider two different aspects of the distribution of the change in Tatsoib: mean change and proportion of a positive change. Evidently they results in different comparisons between the active drug group and the placebo group.

6. Argue that it is reasonable to describe the relation between age and the change in Tatsoib as approximately linear.
Estimate parameters describing the linear relationship. Comment on the interpretation of these estimates.

The relation between age and the change in Tatsoib is approximately linear in the placebo group. The assumption of linearity was accessed using both scatter and residual plots. The mean change in Tatsoib was 0.56 (CI : -.01; 1.12) for individuals of age 45 (the intercept in the model) and the difference in mean increase in Tatsoib between two individuals with a 1 year age difference is 0.12 (CI: 0.08;0.17) when comparing the oldest to the youngest individual (the slope). The standard deviation was 1.63, which expresses that the variation in the change in Tatsoib is ±3.19 for fixed age.

7. Repeat this for group I.

The relation between age and the change in Tatsoib is also approximately linear in the active drug group.

8. Show that the slopes in the two groups can be assumed to be identical.
What is the interpretation of this?

The difference in the slopes is 0.01 (CI: -0.06;.089) comparing the active drug group to the placebo group, which is not statistical different from 0 (p=0.72). We can thus assume that the difference in mean change in Tatsoib between two individuals with a 1 year age difference is the same in the active drug group to the placebo group

9. Do this and write a conclusion on the age-adjusted difference between the two groups.

Using an analysis of covariance we find 1.08 (CI: 0.33;1.83) higher increase in Tatsoib in the active drug group as compared to the placebo group when adjusting for age, which is statistical significant (p=0.005).

10. Comment on the similarities and differences between the analyses and results in Questions 3 and 9.
Discuss the citation on the previous page.

Since there is no difference in age distribution per randomization between the active drug and placebo group the two-sample t-test and analysis of covariance estimates the same mean difference in increase in Tatsoib between the two groups. However, the precision of the estimates of means in the two-sample t-test is determined by the variation in each of the two groups (SD=2.1). In contrast, the precision of the estimates of means in the analysis of covariance is determined by the variation around the lines (SD= 1.6). Since the standard deviation in the analysis of covariance is always smaller than or equal to the standard deviation in the two-sample t-test, we obtain in the former an analysis with higher statistical power and more precise estimates.

The authors correctly conclude that there is no need for adjustment for age in terms of bias, but there is still an advantage of adjustment for age in terms of precision.

11. Do this and write a conclusion on the age-adjusted difference between the two groups. Comment on the similarities and differences between this analysis and the result in Question 10 and what you found in Question 4.

The crude odds ratio for an increase in Tatsoib in the active drug group compared to the placebo group is 3.1 (CI: 1.1;8.4, p=0.030) based on the standard error of the log(OR) of 1.57. The age-adjusted odds ratio of an increase in Tatsoib in the active drug group compared to the placebo group is 6.2 (CI: 1.7;22.1, p= 0.005) based on the standard error of the log(OR) of 4.02.

Conclusion

Two important conclusions, one in terms of difference in estimates between a crude analysis and an adjusted analysis and another in terms of precision of the two estimates. First, the standard definition of a confounder:

(A) risk factor of the outcome,

(B) association with the exposure,

(C) not an intermediate causal factor between the exposure and the outcome

will only ensure that

the crude and adjusted effects are equal if and only if the covariate is **not** a confounder

in nice “linear models” including linear models for the mean (linear regression), models for the risk difference and relative risk. In other “non-linear models” including the logistics regression model and the Cox regression model that will be considered Day 8, the crude and adjusted effects are always different even for a covariate that is **not** a confounder.

Second, in linear regression model one obtained **more** precise estimates when adjusting for a covariate. In most other regression models, like the logistic regression model one obtain **less** precise estimates when adjusting for a covariate.

Do file

```
*******************************************************************
* Solution to Exercise 7-2.
*******************************************************************

cd "D:\Teaching\BasicBiostat\Exercises"

capture log close
log using "solution7-2.log",text replace

use tatsoib.dta,clear

* Q1
histogram age , by(group,legend(off)col(1)) norm name(p1) scale(1.5)
qnorm age if group==0, name(p2) scale(1.5)
qnorm age if group==1, name(p3)scale(1.5)
graph combine  p2 p3 , name(p4) col(1)
graph combine  p1 p4 ,col(2)
graph drop p1 p2 p3 p4
graph export p7_2_1.png,replace

ttest age,by(group)
sdtest age,by(group)

* Q2 + Q3.
histogram ch_tats,by(group,legend(off)col(1)) norm name(p1) scale(1.5)
qnorm ch_tats if group==0, name(p2) scale(1.5)
qnorm ch_tats if group==1, name(p3) scale(1.5)
graph combine  p2 p3 , name(p4) col(1)
graph combine  p1 p4 , col(2)
graph drop p1 p2 p3 p4
graph export p7_2_2.png,replace
sdtest ch_tats , by(group)
ttest ch_tats , by(group)
* A two-sample t-test is identical to a linear regression analysis using
* a categorical independent variable with two levels, here group. Note,
* that the common standard deviation can be read off the output of the
* linear regression.
regress ch_tats i.group
* The command "estimate" allow for the results after a regression analysis
* to be stored and can later be shown and compared to other regression
* results. See later.
estimates store model1

* Q4
* The "if(ch_tats<.)" states that the ch_pos should only be created for
* the non-missing values (missing is defined as "." which Stata consider
* as a very large number).
generate ch_pos=ch_tats>0 if(ch_tats<.)
ci ch_pos if group==1, binom
```

```
ci ch_pos if group==0, binom
cs ch_pos group,or woolf

* Q6.
generate age45=age-45
twoway   ///
   (scatter ch_tats age if group==0, mco(blue))  ///
   (lfit ch_tats age if group==0, lco(blue))  ///
   (scatter ch_tats age if group==1, mco(red))  ///
   (lfit ch_tats age if group==1, lco(red))   ///
    , legend(lab(1 "placebo") lab(2 "") lab(3 "active") lab(4 "") ring(0) pos(11))  ///
    ytit( "change") scale(1.5)
graph export p7_2_6.png,replace

regress ch_tats age45 if group==0
predict res0 if e(sample),res
qnorm res0,scale(1.5) title("Group==0")
graph export p7_2_6b.png,replace

* Q7.
regress ch_tats age45 if group==1
predict res1 if e(sample),res
qnorm res1 ,scale(1.5) title("Group==1")
graph export p7_2_7.png,replace

* Q8
regress ch_tats i.group##c.age45

* Q9
regress ch_tats c.age45 i.group
estimates store model2
predict res if e(sample),res
qnorm res  ,scale(1.5) title("age+group")
graph export p7_2_9a.png,replace

predict fitted if e(sample)
sort age
twoway   ///
   (scatter ch_tats age if group==0, mco(blue))  ///
   (line fitted age if group==0, lco(blue))  ///
   (scatter ch_tats age if group==1, mco(red))  ///
   (line fitted age if group==1, lco(red))  ///
    , legend(lab(1 "placebo") lab(2 "") lab(3 "active") lab(4 "")ring(0) pos(11))  ///
    ytit( "change") scale(1.5)
graph export p7_2_9b.png,replace
* Now, we present the crude and age-adjusted results.
* Option:
* - se: adds standard errors to the display.
estimates table model1 model2, se

* Q11
```

```
logit ch_pos i.group,or
estimates store model1b
logit ch_pos i.group##c.age45,or
logit ch_pos age45 i.group,or
estimates store model2b
estimates table model1b model2b, se

log close
```