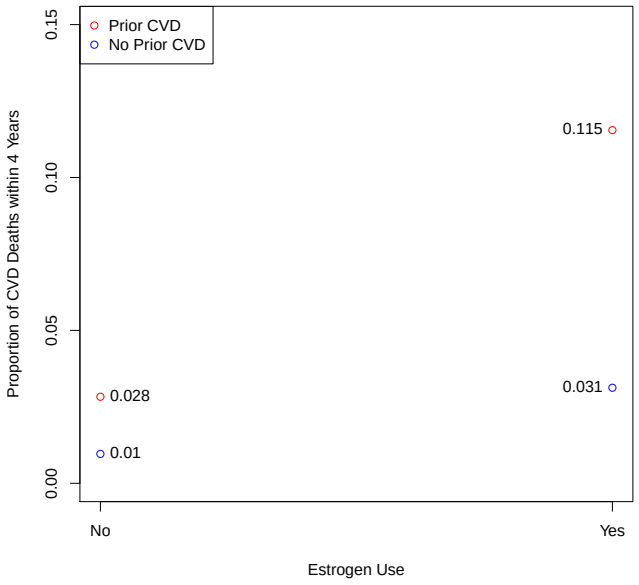
ID #2887
BIOSTAT 536
Homework 2

1. (a) Based on a regression analysis, the estimated difference in risk of CVD mortality between subjects with and without a prior history of estrogen use is 0.038 (SE 0.007), with subjects with a prior history of estrogen use having lower risk of CVD mortality. A 95% confidence interval for this estimate is (0.025, 0.051). Based on a P-value < 0.001, we have sufficient evidence at the 0.05 significance level to reject the null hypothesis of no association between CVD mortality and prior history of estrogen use.

   (b) We are interested in assessing whether there is evidence of effect modification by a history of prior CVD. Since we are using risk difference as our summary measure, we will evaluate the evidence for effect modification by examining the risk difference among people with a history of prior CVD and those without history of prior CVD. If these two risk differences are sufficiently different, we will take this as evidence of effect modification. The following figure will help in assessing whether effect modification is present. Based on this figure, we



   can see that the risk difference for those with prior history of CVD is rather large, while the risk difference for those without prior history of CVD is much smaller. We therefore conclude that there is evidence for effect modification by prior history of CVD.

   (c) Ignoring the possibility of effect modification, we will now examine whether there is evidence

in the dataset that history of prior CVD confounds the relationship between prior estrogen use and CVD mortality. The proportion of people with a prior history of estrogen use among those without prior history of CVD is 0.08. The proportion of people with a prior history of estrogen use among those with a prior history of CVD is 0.02. Based on this, we conclude that history of estrogen use is associated with prior history of CVD in this dataset. As a result, there is evidence that prior history of CVD is a confounder.

(d) After controlling for history of prior CVD, the estimated difference in risk of CVD mortality between subjects with and without a prior history of estrogen use is 0.025 (SE 0.006), with subjects with a prior history of estrogen use having lower risk of CVD mortality. A 95% confidence interval for this estimate is (0.012, 0.038). Based on a P-value < 0.001, we have sufficient evidence at the 0.05 significance level to reject the null hypothesis of no association between CVD mortality and prior history of estrogen use.

(e) To examine whether the estrogen-CVD mortality is further confounded by age after adjusting for prior CVD history, we will examine how including age as a covariate in the regression model used in part (d) affects the estimate of the coefficient associated with estrogen use. If including age as a covariate changes the estimate, we will take this as evidence that age further confounds the estrogen-CVD mortality relationship. The estrogen use coefficient estimate from the model used in part (d) is -0.025, while the coefficient estimate from the model that includes age as a covariate is -0.016. Since the coefficient estimate has changed, we conclude that there is evidence that age further confounds the relationship between CVD mortality and estrogen use.

(f) After adjusting for age and history of prior CVD, the estimated difference in risk of CVD mortality between subjects with and without a prior history of estrogen use is 0.016 (SE 0.006), with those subjects with a prior history of estrogen use having lower risk of CVD mortality. A 95% confidence interval for this estimate is (0.003, 0.029). Based on a P-value of 0.013, we have sufficient evidence at the 0.05 significance level to reject the null hypothesis of no association between CVD mortality and prior history of estrogen use.

2. (a) Based on a logistic regression analysis, the estimated odds ratio is 0.227, with those subjects with a prior history of estrogen use having lower odds of CVD mortality. A 95% confidence interval for this estimate is (0.084, 0.614). Based on a P-value of 0.003, we have sufficient evidence at the 0.05 significance level to reject the null hypothesis of no association between estrogen use and CVD mortality.

(b) We are interested in evaluating the evidence in the data for effect modification by prior history of CVD. Since our measure of association is now the odds ratio, we will examine the odds ratio for subjects with prior history of CVD and the odds ratio for subjects without prior history of CVD. If these two odds ratios are sufficiently different, we will take this as evidence for effect modification by prior history of CVD. The following 2x2 table shows the estimated odds of CVD mortality by estrogen use and prior history of CVD. From this table, we can see that the odds ratio among subjects with prior CVD is 0.247, while the odds ratio among subjects without prior CVD is 0.333. Based on this, there appears to be evidence in the data for effect modification by prior history of CVD.

	Prior Estrogen Use	No Prior Estrogen Use
Prior CVD	0.032	0.131
No Prior CVD	0.010	0.029

(c) Ignoring the possible effect modification, we now examine whether there is evidence in the data of confounding by prior history of CVD. We will examine how including prior history of CVD in the regression model used in part (a) affects the estimate of the coefficient associated with estrogen use. If including prior history of CVD results in a less extreme coefficient estimate, we will take this as evidence that CVD history is a confounder. The estrogen use coefficient estimate from the model used in part (a) is -1.482, while the coefficient estimate from the model that includes CVD history is -1.182. Since including CVD history results in a less extreme coefficient estimate, we take this as evidence that CVD history is a confounder.

(d) After adjusting for history of prior CVD, the estimated odds ratio for CVD mortality among subjects with and without history of estrogen use is 0.307, with those subjects with history of estrogen use having lower odds of CVD mortality. A 95% confidence interval for this estimate is (0.113, 0.835). Based on a P-value of 0.021, we have sufficient evidence to reject the null hypothesis of no association between estrogen use and CVD mortality.

(e) To examine the evidence in the dataset that the prior disease adjusted analysis of an association between estrogen-CVD mortality is further confounded by age, we will examine whether including age as a covariate in our regression model from part (d) results in a less extreme estimate for the coefficient associated with estrogen. If the coefficient changes, this will be taken as evidence that age is further confounding the association between estrogen-CVD mortality after adjusting for prior disease. When age is included in the model, the estimate for the coefficient associated with estrogen use is -1.017, while the coefficient estimate for estrogen use from the model that only adjusts for previous CVD is -1.182. Since adjusting for age results in a less extreme estimate, we can conclude that there is evidence that age is a confounder.

(f) After adjusting for history of prior CVD and age, the estimated odds ratio for CVD mortality between subjects with and without history of estrogen use is 0.362, with those subjects with history of estrogen use having lower odds of CVD mortality. A 95% confidence interval for this estimate is (0.131, 0.999). Based on a P-value of 0.050, we have sufficient evidence at the 0.05 significance level to reject the null hypothesis of no association between estrogen use and CVD mortality after adjustment for CVD history and age.

3. (a) Based on a regression analysis, the estimated risk ratio for CVD mortality between subjects with and without history of estrogen use is 0.236, with those subjects with prior history of estrogen use estimated to have lower risk of CVD mortality. A 95% confidence interval for this estimate is (0.090, 0.631). Based on a P-value of 0.004, we have sufficient evidence at the 0.05 significance level to reject the null hypothesis of no association between CVD mortality and estrogen use.

(b) We are interested in assessing the evidence for effect modification by previous history of CVD

mortality. Since we are using the risk ratio as our measure of association, we will examine the risk ratio among those with a previous history of CVD and the risk ratio among those without a previous history of CVD. If these two risk ratios are sufficiently different, we will take this as evidence of effect modification by previous history of CVD. The risk ratio among those without a history of CVD is 0.340, while the risk ratio among those with a history of CVD is 0.271.

(c) To examine whether previous history of CVD confounds the relationship between estrogen and CVD mortality, we will examine how including CVD history as a covariate in the model used in part (a) affects the coefficient estimate associated with estrogen use. If including CVD history in the model changes the coefficient estimate, we will take this as evidence that CVD history is a confounder. The estrogen use coefficient estimate from the model used in part (a) is -1.443, while the coefficient estimate from the model that includes CVD history is -1.144. Since these two estimates are different, we conclude that there is evidence that CVD history is a confounder.

(d) After adjusting for previous history of CVD, the estimated risk ratio between subjects with and without a history of estrogen use is 0.320, with subjects with a prior history of estrogen use having lower risk of CVD mortality. A 95% confidence interval for this estimate is (0.120, 0.851). Based on a P-value of 0.022, there is sufficient evidence at the 0.05 significance level to reject the null hypothesis no association between estrogen use and CVD mortality after adjusting for previous history of CVD.

(e) To examine whether the estrogen-CVD mortality association is further confounded by age even after adjusting for CVD, we will examine how including age as a covariate in the regression model used in part (d) affects the estimate of the coefficient associated with estrogen use. If the estimate varies between the two models, we will take this as evidence that age is a confounder. The estrogen use coefficient estimate from the model used in part (d) is -1.144, while the coefficient estimate from the model that includes age as a covariate is -1.000. Since these two estimates differ, we take this as evidence that age is a confounder.

(f) After adjusting for previous history of CVD and age, the estimated risk ratio between subjects with and without a history of estrogen use is 0.368, with those subjects with a prior history of estrogen use having lower risk of CVD mortality. A 95% confidence interval for this estimate is (0.137, 0.986). Based on a P-value of 0.047, we have sufficient evidence at the 0.05 significance level to reject the null hypothesis of no association between estrogen use and CVD mortality after adjustment for previous CVD history and age.

4. Regardless of the measure of association used, all three approaches came to similar conclusions regarding potential confounders and inference about the association between estrogen and CVD mortality.

   One advantage of the approach using the difference in risks is that it easily interpreted. A drawback, however, is that linear regression does not take into account the fact that risks are constrained to be between 0 and 1. This is especially problematic when the estimated risks are likely to be close to these boundary values.

One advantage of the approach using odds ratios is that logistic regression does place the necessary restraints on the estimated probabilities. However, the odds ratio is difficult to interpret.

One advantage of the approach using the risk ratio is that it is more easily interpreted than the odds ratio. However, one drawback is that we cannot estimate the absolute risk.