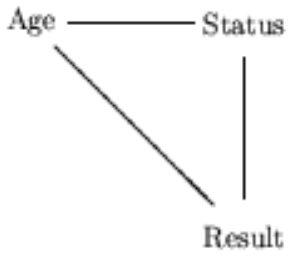
1. (a) The line for Age:Result is testing to see if Age:Result should be added to the model that contains the 3 main effects (Age, Status, and Result) and the Age:Status interaction. It is testing the hypothesis that there is no Age:Result interaction given that Age, Status, Result and Age:Status are in the model.

   (b) No it is not the maximal model. For a contingency table the maximal model would contain all the main effects and all possible interactions. The model selected by S-plus does not contain the 3-way interaction (Age:Status:Result). Note, that the Residual Deviance for the last line of the ANOVA table is not 0, therefore we know this model is not maximal.

   (c) The association graph is:



   All factors are directly related to each other.

   (d) The contingency table cannot be collapsed. A 3-way table can only be collapsed on a factor that is not directly related to both other factors. In this case all factors are directly related to both other factors.

   (e) It makes sense to treat Breathing Test Result as the response since it is possible that changes in Age or Smoking Status will cause a change in Breathing Test Result. It does not make sense to say that a change in Breathing Test Result will cause a change in either of the other two variables.

   (f) Create a table that gives the Breathing Test Result as percents (or proportions) for each combination of Age and Status:

		Breathing Test Result		
Age	Smoking Status	Normal	Borderline	Abnormal
<40	Never Smoked	95.1%	4.4%	0.5%
	Former Smoker	91.4%	7.1%	1.4%
	Current Smoker	92.3%	6.2%	1.5%
40-59	Never Smoked	92.1%	5.6%	2.2%
	Former Smoker	86.8%	9.0%	4.2%
	Current Smoker	78.8%	12.5%	8.7%

   From this table we see that for each level of Smoking Status, the percentage of workers classified as having an Abnormal or Borderline Breathing Test result is higher for the $40-50$ age group than for the < 40 age group.

For the < 40 age group, the Former Smoker group and the Current smoker group had approximately the same percentages in each of the Breathing Test categories. Both these groups have a higher percentage of Abnormal and Borderline results than the Never Smoker group.

For the $40 - 50$ age group, the Current Smoker group has higher percentages of Abnormal and Borderline results than the Former Smoker group which in turn has higher percentages of Abnormal and Borderline results than the Never Smoker group.

2. (a) Relative risk$= 0.00140/0.00010 = 14$. This means that the probability a smoker dies from lung cancer is $14\times$ the probability a non-smoker dies from lung cancer.

   (b) The odds ratio of dying from heart disease for smokers relative to non-smokers is:

$$\frac{0.00669/(1-0.00669)}{0.000413/(1-0.000413)} = 1.624$$

   This odds ratio will be approximately equal to relative risk since the probability of dying from heart disease is quite small for both smokers and non-smokers.

   (c) We need to compare the the difference in the probability of dying from lung cancer between smokers and non-smokers $(0.0014 - 0.00010 = 0.00013)$ with the difference in the probability of dying from heart disease between smokers and non-smokers $(0.00669 - 0.00413 = 0.00256)$. Since this difference is larger for heart disease, the reduction in deaths due to heart disease would be more than the reduction in deaths due to lung cancer.

3. (a) For case control studies we cannot estimate the probability of having the disease (given a risk factor) but we can estimate the odds ratio of having the disease for a risk factor.

   (b) The estimated odds ratio is:

$$\frac{73 * 82}{21 * 188} = 1.516211$$

   This indicates that the odds a non-smoking woman who is married to a smoker has lung cancer is 1.5 times the odds a non-smoking woman who is married to a non-smoker has lung cancer.

   (c) First find a 95% confidence interval for the log odds.

$$\log(1.516) \pm 1.96 \times \sqrt{\frac{1}{73} + \frac{1}{82} + \frac{1}{21} + \frac{1}{188}} = (-0.13, 0.97)$$

   Then convert to an interval for the odds ratio:

$$(\exp(-0.13), \exp(0.97)) = (0.87, 2.63)$$

   (d) The 95% confidence interval found in (c) contains 1 (i.e. odds for women married to smokers are the same as the odds of women married to non.smokers) so there will not be strong evidence of a connection between passive smoking and lung cancer (the P-value will be > 0.05). However, since 1 is close to the lower boundary of the confidence interval, the P-value won't be much bigger than 0.05 so there will be weak evidence of a connection.