

Data analysis and interpretation - Assignment solutions

Part I

1. (a) Restriction

The methods for this paper were included in the Sources of Bias assignment, so we need to look at that paper to find the answer. The 3rd paragraph of the Methods section (page 118) says: "One hundred and fifty-six respondents reported having been hospitalized for MI before their menopause For each of these case, we selected 20 control subjects from respondents ... and who were premenopausal at the time of hospitalization of the case." Thus, it appears that both cases and controls were premenopausal at least to the time that the MI occurred or the comparable date for the matched controls. The controls were matched on several factors, including being premenopausal. But because ONLY premenopausal women were studied, the method is Restriction (to one level of the variable) rather than matching (enforcing the same distribution of the matching variable) and stratified analysis (which involves dividing the dataset into strata, not collecting data from only one stratum).

2.

	OC use	
	Current	Never
MI	$(16 + 7) = 23$	$(42 + 53) = 95$
No MI	$(190 + 114) = 304$	$(991 + 1045) = 2036$

3.

$$\text{OR} = \frac{(23)(2036)}{(95)(304)} = 1.62$$

4. Age (at hospitalization) is not a confounder:

Crude OR = 1.6

Stratum-specific OR's are 2.0 and 1.2, so that the crude lies well within their range.

5. The concept of effect modification can be approached from different perspectives. One perspective is to regard effect modification as a departure from a multiplicative model, since the multiplicative model is most often employed in investigations of etiology and, from a practical standpoint, departure from multiplicativity means that a weighted average of stratum-specific ratio measures of effect (e.g., OR's) may be misleading. This example is complicated by the fact that controls were matched to cases on age, so that the effect of age cannot be evaluated from the data presented in the paper. However, homogeneity across strata of age can be examined. We already know, of course, that MI rates increase sharply with older age.

If the OR for Current OC use were the same in the two age strata, then we could conclude that the observed odds ratios fit a multiplicative model, so that there would be no effect modification based on the above perspective. However, the OR_{OC} for the older women is smaller than the OR_{OC} for the younger women. That suggests that the combined effect of current OC use and greater age is less than would be expected based on a multiplicative model, which could be interpreted as evidence of effect modification. The evidence is weak, however, since although confidence intervals are not presented, the OR estimates are based on rather small numbers of exposed cases and are therefore imprecise. Unless a statistical test for heterogeneity of the OR across strata indicated that the observed difference in the OR's (1.2 versus 2.0) is beyond that expected from chance alone, one would say that there is, at most, slight evidence for effect modification.

The other perspective on effect modification relates to impact, i.e., that if the combined effect is greater than expected from an additive model, then interventions may be worth targeting to those dually exposed. This perspective cannot be fully investigated in the data we have here, because of the matching. But since the combined effect is less than expected based on a multiplicative model, the combined effect is presumably not much greater than expected based on an additive model.

6. (a) Yes: OR for CIG only is 5.0 (2.7-9.0).

(b) Greater than: OR for CIG + HYP = 8.9, compared to OR for HYP only = 7.6. But the confidence intervals are broad and have substantial overlap, so "cannot determine" is also a reasonable conclusion.

7. Hospitalization for MI and OC use in nurses with no history of hypertension.

	OC use	
	Current	Never
MI	(5 + 7) = 12	(12 + 39) = 51
No MI	(150 + 107) = 257	(1022 + 669) = 1691
OR	$= \frac{(12)(1691)}{(51)(257)} = 1.55$	

8. From last line of the table: 170 (31 - 1100)

9. $OR_{CIG,HYP,OC}$ [multiplicative model]

$$\begin{aligned} &= OR_{CIG|HYP,OC} \times OR_{HYP|CIG,OC} \times OR_{OC|CIG,HYP} \\ &= 5.0 \times 7.6 \times 2.8 \\ &= 106 \end{aligned}$$

(This is less than the observed OR.)

Part II

1. Past OC use (Relative risk estimate 0.9) and Overweight (Relative risk estimate 1.2) are both not importantly related to MI risk. Though an OR of 1.2 indicates some elevation of risk, the confidence interval extends so far below 1.0 that the elevation is consistent with an interpretation in terms of chance.

2. The relationship between the coefficient for Current OC use and the relative risk estimate is:

$$\text{Relative risk estimate} = OR = \exp(0.59) = e^{0.59} = 1.8$$

3. Age has not been controlled in this logistic model. The cases and controls were, however, matched by year of birth. It is not clear that this matching eliminates possible confounding by age. Nevertheless, from Table 1, there does not appear to be confounding by age, and while it is theoretically possible to have confounding in the multivariable analysis even though none was observed in the stratified analysis of Table 1, that likelihood is probably small.

4. The crude OR from Table 1 is 1.6; the summary OR (controlling for age) is also 1.6. there may be a small amount of confounding caused by the other risk factors, therefore, since the OR from the multiple logistic model is 1.8 for Current OC use. But the difference between 1.6 and 1.8 is not important.

5. This logistic model consists entirely of indicator (dichotomous) variables. In part, this fact was necessitated by the study questionnaire, which asked for history of various conditions, rather than their actual values (e.g., blood pressure). Overweight could presumably have been entered as a continuous variable. Using a single indicator variable to express the value of a continuous variable loses information. There are some offsetting advantages, however.

6. There are no interaction (product) terms in this model, so no provision has been made for deviation from the underlying model that the odds ratio for a combination of factors equals the product of their respective odds ratios (or equivalently, that the logarithm of the odds of MI equals the sum of the logarithms of the odds ratios for the factors, plus a constant), and that this relationship is not altered by the value of other variables in the model.
7. The odds ratio for the combination of cigarette smoking and Current OC use (compared to neither factor) is, since there are no product terms to consider, simply the product of the odds ratios (relative risks) for each of these two factors:

$$OR_{\text{cig,OC}} = OR_{\text{cig}} \times OR_{\text{OC}} = 2.8 \times 1.8 = 5.04 \approx 5.0$$

This OR is close to the value in Table 4 for "OC and CIG only" (5.6) for normotensive individuals. For hypertensives, the estimate for the joint effect of OCs and cigarettes is obtained by dividing 170 (the OR for "OC, CIG, and HYP") by 7.6 (the OR for "HYP only"); among hypertensives, therefore, the OR for OCs combined with cigarette smoking is $170 / 7.6 = 22.4$, a value well above the 5.0 estimate from the model. Since there are no product terms involving hypertension, the logistic model assumes that the OR for each factor or combination of factors is unaffected by hypertension. In other words, the OR from this logistic model represents a type of average of the OR in normotensives and that in hypertensives.

The difference between the value from the logistic model (5.0) and the values from the stratified analysis (5.6 and 22.4) can be attributed to the "smoothing effect" of the logistic regression, which forces all odds ratios to fit the form of the assumed model (of multiplicative odds ratios with no heterogeneity). From the figures in Table 4, it is clear that most of the cases and controls were not hypertensive, so the logistic model odds ratio estimates will primarily reflect the odds ratios in normotensives. Hence the logistic value is closer to the 5.6 than to the 22.4.

Another possible reason for the difference between the stratified and logistic regression odds ratios is that the latter control for a variety of other risk factors that are not included in the stratified analysis. If these other factors confound the OC and cigarette smoking relationship with MI risk, then the stratified analysis results in Table 4 may be confounded.