# Chapter 14
# Dealing with confounding in the analysis

In the previous chapter we discussed briefly how confounding could be dealt with at both the design stage of a study and during the analysis of the results. We then mentioned that there are two main statistical procedures that we can use in the analysis: *stratification* and *regression modelling*. In this chapter we will discuss these two approaches in more detail. Obviously, these techniques can be applied only if data on potential confounding factors are available. Thus, potential confounding variables have to be identified at the design stage of the study to ensure that valid information on them is collected.

## 14.1 Introduction to stratification

A confounding factor is one that is related to both the exposure and the outcome variables and that does not lie on the causal pathway between them (see Section 13.2). Ignoring confounding when assessing the association between an exposure and an outcome variable can lead to an overestimate or underestimate of the true association between exposure and outcome and can even change the direction of the observed effect.

***Example 14.1.*** *In a hypothetical case–control study to examine the relationship between smoking and ovarian cancer among nulliparous women, the results shown in Table 14.1 were obtained.*

| | Smoking | | Total |
|---|---|---|---|
| | Yes | No | |
| Ovarian cancer cases | 24 (*a*) | 36 (*b*) | 60 ($n_1$) |
| Controls | 58 (*c*) | 40 (*d*) | 98 ($n_0$) |
| **Total** | **82 ($m_1$)** | **76 ($m_0$)** | **158 (*N*)** |
| Crude odds ratio = (24/36) / (58/40)=0.46 | | | |
| 95% confidence interval = 0.23–0.93 | | | |
| $\chi^2$ = 5.45 on 1d.f.; $P$ = 0.02 | | | |

**Table 14.1.**
Results of a case–control study on smoking and ovarian cancer: hypothetical data.

In Example 14.1, women with ovarian cancer had a much lower prevalence of smoking (24/60 = 40%) compared with the controls (58/98 = 59%). This suggests that smoking protects against ovarian cancer (odds ratio (OR) = 0.46). As discussed in the previous chapter, there are several possible explanations for this finding:

(i) *Bias*: the observed odds ratio of 0.46 does not accurately represent the true odds ratio because of either selection or measurement bias.

(ii) *Chance*: the observed association between smoking and ovarian cancer arose by chance. The 95% confidence interval around the observed odds ratio is equal to 0.23–0.93 and the $\chi^2$ test yields $P$=0.02. Thus, chance is an unlikely explanation for the finding.

(iii) *Confounding*: the observed odds ratio of 0.46 is due to the effect of another variable. For example, it may be that women who smoked were different in other respects from non-smokers and less likely to develop ovarian cancer because of this, rather than because of smoking.

(iv) *Causation*: smoking reduces the risk of ovarian cancer and the 95% confidence interval indicates how precisely the sample estimate corresponds to the true effect in the population.

In Example 14.1, it is possible that the association between smoking and ovarian cancer arose because of the confounding effect of other factors such as oral contraceptive use. The results shown in Table 14.1 are for all women combined regardless of their history of oral contraceptive use. To assess whether oral contraceptive use is a confounder, we need to look at the association between smoking and ovarian cancer separately for oral contraceptive users and never-users. This is shown in Table 14.2.

**Table 14.2.**
Hypothetical case–control study on smoking and ovarian cancer described in Example 14.1: results presented separately for never-users and ever-users of oral contraceptives.

**Never-users of oral contraceptives**

| | Smoking | | Total |
|---|---|---|---|
| | **Yes** | **No** | |
| Ovarian cancer cases | 9 ($a_1$) | 32 ($b_1$) | 41 ($n_{11}$) |
| Controls | 8 ($c_1$) | 28 ($d_1$) | 36 ($n_{01}$) |
| **Total** | **17 ($m_{11}$)** | **60 ($m_{01}$)** | **77 ($N_1$)** |

Odds ratio = 0.98 (95% confidence interval = 0.30–3.37)

**Ever-users of oral contraceptives**

| | Smoking | | Total |
|---|---|---|---|
| | **Yes** | **No** | |
| Ovarian cancer cases | 15 ($a_2$) | 4 ($b_2$) | 19 ($n_{12}$) |
| Controls | 50 ($c_2$) | 12 ($d_2$) | 62 ($n_{02}$) |
| **Total** | **65 ($m_{12}$)** | **16 ($m_{02}$)** | **81 ($N_2$)** |

Odds ratio = 0.90 (95% confidence interval = 0.23–4.40)

In each category (or *stratum*) of oral contraceptive use, the prevalence of smoking was similar in women with and without ovarian cancer (22% versus 22% among never-users and 79% versus 81% among users). However, when both oral contraceptive users and never-users were combined (Table 14.1), there was a marked difference in the prevalence of smoking between cases and controls (40% versus 59%). Two factors were responsible for this finding:

1. Among the controls, smokers had a much higher prevalence of oral contraceptive use (50/58 = 86%) than non-smokers (12/40 = 30%), that is there was an association between these two variables (Table 14.3).

| | | Oral contraceptive use | | Total |
|---|---|---|---|---|
| | | Ever | Never | |
| Smoking | Yes | 50 | 8 | 58 |
| | No | 12 | 28 | 40 |
| | **Total** | **62** | **36** | **98** |

**Table 14.3.** Hypothetical case–control study described in Example 14.1: distribution of controls by smoking habits and oral contraceptive use.

Note that the association between smoking and oral contraceptive use was examined among the controls rather than the cases, or the two groups taken together. This is because controls should represent the population from which the cases were drawn and we need to assess that association in the general population. In a cohort or intervention study, the association would be looked at by constructing a similar table, replacing the number of controls with person-years at risk if the appropriate measure of effect was a rate ratio or numbers of persons at risk at the start of the follow-up if the measure of effect was a risk ratio.

2. Oral contraceptive use is considerably lower among ovarian cancer cases than among controls. The data from Table 14.2 can be rearranged so that smoking is ignored (Table 14.4).

| | Oral contraceptive use | | Total |
|---|---|---|---|
| | Ever | Never | |
| Ovarian cancer cases | 19 | 41 | 60 |
| Controls | 62 | 36 | 98 |
| **Total** | **81** | **77** | **158** |

**Table 14.4.** Hypothetical case–control study described in Example 14.1: distribution of cases and controls by oral contraceptive use.

Only 32% (=19/60) of the women with ovarian cancer were oral contraceptive users, whereas 63% (=62/98) of the controls were users.

Since oral contraceptive use in these data was associated with *both* the exposure (smoking) and the outcome of interest (ovarian cancer), it acted as a confounding factor. As a result, when the data for both users and never-users were combined, the result suggested an association between smoking and ovarian cancer far stronger than really existed (*positive confounding*). In other situations (as in Example 14.2; see next section), combining strata in the presence of a confounder may mask an effect that really exists (*negative confounding*), or even show an effect in the opposite direction to the true one.

## 14.2 The Mantel–Haenszel summary measures of effect

When we analyse the results of a study to look for evidence of a particular exposure–outcome association, we usually start by including in the analy-

sis all the subjects in our study sample. This analysis provides a crude estimate of the effect of the exposure on the outcome of interest. The next logical step is to divide our study sample into several subgroups or strata, defined by a potential confounding variable, to see if the results are consistent across the strata. This approach is very informative, as it describes how the effect of the exposure on the outcome of interest varies across subgroups of subjects with different characteristics. We can simply report the stratum-specific effect estimates and their confidence intervals. Each of these stratum-specific estimates is supposed to be homogeneous in relation to the potential confounding variable and therefore they are unconfounded.

Usually, however, we are not much interested in the stratum-specific results *per se* and would rather have a single overall estimate. In other words, we would like to be able to calculate a summary effect estimate which, in contrast to the *crude* estimate, would take into account the confounding effect of the stratifying variable. Such *adjusted* estimates can be calculated by *pooling* results across strata. But even if the true effect is the same in all strata, we would expect our estimates to differ because of random variation. Pooling takes this into account by giving greater weight to effect estimates from larger strata. It involves calculating a *weighted average* of the individual stratum-specific estimates by choosing a set of weights that maximizes the statistical precision of the adjusted effect estimate. There are several alternative weighting procedures which achieve precise summary effect estimates. In this section, we concentrate on a procedure derived by Mantel and Haenszel which is widely used and relatively simple to apply.

### 14.2.1 Mantel–Haenszel odds ratio

Let us consider again Example 14.1. Since oral contraceptive use is a confounder of the relationship between smoking and ovarian cancer, we should not combine the data from ever-users and never-users for the analysis. Thus, the crude odds ratio of 0.46 obtained from Table 14.1 is not appropriate. We could just calculate separate odds ratios and their 95% confidence intervals for each group of oral contraceptive users, as shown in Table 14.2. But we would like to be able to summarize the overall results of the study in a way that removes the confounding effect of oral contraceptive use. The Mantel–Haenszel odds ratio, denoted $OR_{MH}$, gives a *weighted average of the odds ratios in the different strata*, where those from larger strata are given more weight.

To calculate the $OR_{MH}$, we start by constructing $2 \times 2$ tables of exposure by outcome for the separate strata of the confounder, as illustrated in Table 14.2. The $OR_{MH}$ can then be obtained by applying the following formula:

$$OR_{MH} = \frac{a_1 d_1/N_1 + a_2 d_2/N_2}{b_1 c_1/N_1 + b_2 c_2/N_2}$$

If we apply this formula to the data from our ovarian cancer case–control study (Table 14.2), we obtain

$$OR_{MH} = \frac{(9 \times 28)/77 + (15 \times 12)/81}{(32 \times 8)/77 + (4 \times 50)/81} = \frac{5.49}{5.79} = 0.95$$

Thus, the *odds ratio for smoking adjusted for oral contraceptive use* is 0.95. This adjusted odds ratio contrasts with the *crude odds ratio* of 0.46 obtained from Table 14.1. Adjusting for oral contraceptive use gives an odds ratio much closer to unity, which means that the protection afforded by smoking, if any, is far less strong than the initial result led us to think.

The above formula can easily be extended to more than two strata, by summing both the numerator and the denominator over all strata:

$$OR_{MH} = \frac{\Sigma a_i d_i / N_i}{\Sigma b_i c_i / N_i}$$

In this formula, Σ means sum and the subscript *i* stands for the subscripts 1, 2, 3, ..., which represent each of the strata.

We can calculate a confidence interval around the $OR_{MH}$ and a Mantel–Haenszel $\chi^2$ test by using the formulae given in Appendix 14.1, Section A14.1.1. In our ovarian cancer example, the 95% confidence interval is 0.42–2.16. The $\chi^2$ is equal to 0.016 with one degree of freedom, which corresponds to $P = 0.93$. Thus, after adjusting for oral contraceptive use, the effect of smoking on ovarian cancer is no longer 'statistically significant'.

### 14.2.2 Mantel–Haenszel risk ratio

The Mantel–Haenszel method can also be used to obtain an *adjusted risk ratio*. In Example 14.2, the crude analysis shows that workers exposed to the particular chemical substance had a 52% higher risk of developing lung cancer than those not exposed (Table 14.5). Before concluding that the chemical substance is associated with an increased risk of lung cancer, we need to exclude the possibility that smoking, rather than the occupational exposure, is the explanation for the observed association. To do this, we need to examine the data separately for smokers and non-smokers.

Table 14.6 shows that the stratum-specific risk ratios are higher than the crude risk ratio (2.0 versus 1.52). This is an example of *negative confounding*. It arose because the prevalence of smoking, an important risk factor for lung cancer, was much lower among workers exposed to the chemical substance (4000/84 000=5%) than among those not exposed (16 000/96 000=17%) (Table 14.7).

***Example 14.2.*** *Suppose that a cohort study was set up to investigate whether occupational exposure to a particular chemical substance was associated with an increased risk of lung cancer. All study subjects were followed up for five years after entry into the study (or until diagnosis of lung cancer if earlier). The results of this study are shown in Table 14.5.*

**Table 14.5.**
Results from a cohort study on occupational exposure to a particular chemical substance and lung cancer: hypothetical data.

| | | Exposure to chemical substance | | Total |
|---|---|---|---|---|
| | | Yes | No | |
| **Lung cancer** | Yes | 480 ($a$) | 360 ($b$) | 840 ($n_1$) |
| | No | 83 520 ($c$) | 95 640 ($d$) | 179 160 ($n_0$) |
| | **Total** | **84 000 ($m_1$)** | **96 000 ($m_0$)** | **180 000 ($N$)** |

Crude risk ratio = (480/84 000)/(360/96 000) = 5.71 per 1000/3.75 per 1000 = 1.52
95% confidence interval = 1.33–1.75
$\chi^2$ = 37.21 on 1d.f.; $P < 0.0001$

**Table 14.6.**
Hypothetical cohort study on occupational exposure to a particular chemical substance and lung cancer described in Example 14.2: results presented separately for smokers and non-smokers.

**Smokers**

| | | Exposure to chemical substance | | Total |
|---|---|---|---|---|
| | | Yes | No | |
| **Lung cancer** | Yes | 80 ($a_1$) | 160 ($b_1$) | 240 ($n_{11}$) |
| | No | 3920 ($c_1$) | 15 840 ($d_1$) | 19 760 ($n_{01}$) |
| | **Total** | **4000 ($m_{11}$)** | **16 000 ($m_{01}$)** | **20 000 ($N_1$)** |

Risk ratio = (80/4000)/(160/16 000) = 20 per 1000 / 10 per 1000 = 2.0
95% confidence interval = 1.53–2.61

**Non-smokers**

| | | Exposure to chemical substance | | Total |
|---|---|---|---|---|
| | | Yes | No | |
| **Lung cancer** | Yes | 400 ($a_2$) | 200 ($b_2$) | 600 ($n_{12}$) |
| | No | 79 600 ($c_2$) | 79 800 ($d_2$) | 159 400 ($n_{02}$) |
| | **Total** | **80 000 ($m_{12}$)** | **80 000 ($m_{02}$)** | **160 000 ($N_2$)** |

Risk ratio = (400/80 000)/(200/80 000) = 5.0 per 1000 / 2.5 per 1000 = 2.0
95% confidence interval = 1.69–2.37

**Table 14.7.**
Hypothetical cohort study described in Example 14.2: distribution of study subjects by occupational exposure and smoking habits.

| | | Exposure to chemical substance | | Total |
|---|---|---|---|---|
| | | Yes | No | |
| **Smoking** | Yes | 4000 | 16 000 | 20 000 |
| | No | 80 000 | 80 000 | 160 000 |
| | **Total** | **84 000** | **96 000** | **180 000** |

We can obtain a Mantel–Haenszel summary estimate of the common risk ratio ($R_{MH}$) by applying the following formula to our data

$$R_{MH} = \frac{\Sigma a_i m_{0i}/N_i}{\Sigma b_i m_{1i}/N_i}$$

Thus,

$$R_{MH} = \frac{(80 \times 16\,000)/20\,000 + (400 \times 80\,000)/160\,000}{(160 \times 4\,000)/20\,000 + (200 \times 80\,000)/160\,000} = \frac{264}{132} = 2.0$$

The calculation of confidence intervals around $R_{MH}$ and of the Mantel-Haenszel $\chi^2$ is presented in Section A14.1.2. In our example, the 95% confidence interval is 1.73 to 2.30. The $\chi^2$ is equal to 92.99 with one degree of freedom, which gives $P < 0.0001$. Thus, there is strong evidence that the occupational exposure was associated with an increased risk of lung cancer and this effect was even stronger when differences in smoking habits between exposed and unexposed workers were taken into account.

### 14.2.3 Mantel–Haenszel rate ratio

The Mantel–Haenszel method can also be applied when the appropriate measure of effect is the rate ratio rather than the risk ratio. It gives an *adjusted rate ratio* (denoted $RR_{MH}$) by calculating a weighted average of the rate ratios in the different strata.

***Example 14.3.*** *Suppose that a cohort of healthy women aged 45–50 years was followed up to examine risk factors for various female cancers. At the time of entry into the study, the women completed a questionnaire on sociodemographic variables and gynaecological and reproductive history. A total of 1141 cervical cancer cases occurred during the follow-up period. The relationship between cervical cancer and having ever had a Pap smear test (as reported in the initial questionnaire) is shown in Table 14.8.*

| | Pap smear | | Total |
|---|---|---|---|
| | Ever | Never | |
| Cases | 17 ($a$) | 1124 ($b$) | 1141 ($n$) |
| Person-years at risk | 71 184 ($y_1$) | 1 518 701 ($y_0$) | 1 589 885 ($y$) |
| Rate per 100 000 pyrs | 23.9 ($r_1$) | 74.0 ($r_0$) | 71.8 ($r$) |
| Crude rate ratio = (17/71 184)/(1 124/1 518 701) = 0.32 | | | |
| 95% confidence interval = 0.20 – 0.52 | | | |
| $\chi^2 = 23.69$ on 1d.f.; $P < 0.001$ | | | |

**Table 14.8.**
Results from a cohort study on Pap smear testing and cervical cancer: hypothetical data.

In Example 14.3, the crude rate ratio and its confidence interval are consistent with a decrease in the incidence of cervical cancer among women who reported in the initial questionnaire having ever had a Pap smear test. Other studies have shown that there is a socioeconomic gradient in the incidence of cervical cancer, with women from poor socioeconomic backgrounds being at higher risk. Thus, socioeconomic factors might have confounded the association between Pap smear testing and cervical cancer if, for instance, women from a high social background were more likely to visit their doctors and had a Pap smear as part of their regular medical examination. To clarify this issue, we need to examine the relationship between Pap smear testing and cervical cancer separately for women from different socioeconomic backgrounds. This is shown in Table 14.9, where a woman's educational level is used as a marker of socioeconomic status.

**Table 14.9.**
Hypothetical cohort study on Pap smear testing and cervical cancer described in Example 14.3: results stratified by women's educational level.

**High educational level**

| | Pap smear | | Total |
|---|---|---|---|
| | Ever | Never | |
| Cases | 13 ($a_1$) | 697 ($b_1$) | 710 ($n_1$) |
| Person-years at risk | 38 346 ($y_{11}$) | 828 149 ($y_{01}$) | 866 495 ($y_1$) |
| Rate per 100 000 pyrs | 33.9 ($r_{11}$) | 84.2 ($r_{01}$) | 81.9 ($r_1$) |
| Rate ratio = 0.40; 95% confidence interval = 0.23–0.69 | | | |

**Low educational level**

| | Pap smear | | Total |
|---|---|---|---|
| | Ever | Never | |
| Cases | 4 ($a_2$) | 427 ($b_2$) | 431 ($n_2$) |
| Person-years at risk | 32 838 ($y_{12}$) | 690 552 ($y_{02}$) | 723 390 ($y_2$) |
| Rate per 100 000 pyrs | 12.2 ($r_{12}$) | 61.8 ($r_{02}$) | 59.6 ($r_2$) |
| Rate ratio = 0.20; 95% confidence interval = 0.08–0.54 | | | |

The formula for the Mantel–Haenszel summary estimate of the common rate ratio is

$$RR_{MH} = \frac{\Sigma a_i y_{0i}/y_i}{\Sigma b_i y_{1i}/y_i}$$

Thus, in our example,

$$RR_{MH} = \frac{(13 \times 828\,149)/866\,495 + (4 \times 690\,552)/723\,390}{(697 \times 38\,346)/866\,495 + (427 \times 32\,838)/723\,390} = \frac{16.24}{50.23} = 0.32$$

Thus, educational level was not a confounder of the effect of Pap smear on cervical cancer in these data, since the crude and the adjusted rate ratios have exactly the same value (0.32).

By applying the formulae given in the Appendix (see Section A14.1.3), we obtain a Mantel–Haenszel $\chi^2$ of 23.73 with one degree of freedom, corresponding to $P < 0.001$. The 95% confidence interval is 0.20 to 0.52. Thus, chance is an unlikely explanation of this finding.

Note that the Mantel–Haenszel method of controlling for confounding is similar to the method of standardization used to calculate age-adjusted rates in Chapter 4. Both these methods are referred to as *stratified analyses*, because we look at an exposure by a response for the different *strata* (levels) of a confounder. They differ, however, in the set of weights used to calculate the weighted average of the rate ratios in the different strata (see Section 4.3.3).

## 14.3 How to identify confounders

In order to be able to examine the effect of potential confounding variables in an analysis, we need to identify them at the design stage of the study. This should be done by taking into account findings from previous epidemiological studies and what is known about the possible etiological mechanisms of the disease under study. Age and gender are obvious potential confounders in practically all studies. Smoking will also be a potential confounder in any study examining the relationship between a particular exposure and lung cancer. It would be necessary to exclude the possibility that smoking rather than the exposure under study is responsible for any association that may be found. Potential confounding variables should also include factors such as socioeconomic status or place of residence, which are just proxy measures of more direct but unknown causes of disease.

Not all factors suspected of being potential confounding factors will actually lead to confounding of the exposure–outcome relationship in a particular study. But how do we know if a particular variable really is a confounder? We defined a confounder as a factor that is associated with both exposure and disease (and is not on the causal pathway). However, this may be difficult to assess in practice. For instance, with a large sample, small but statistically significant associations could be found between the confounder and the exposure, and between the confounder and the disease; however, they may not be strong enough to lead to confounding. Thus, the presence of or absence of confounding should not be assessed by using a statistical test of significance. The magnitude of confounding in any study should be evaluated by observing the degree of discrepancy between the crude and adjusted estimates. If there is no difference between these two estimates, the observed exposure–outcome effect was not confounded by the potential confounding variable. A large difference, as seen in Example 14.1, indicates the presence of confounding and implies that the *adjusted* summary measure is a better estimate of the effect of the exposure on the outcome of interest than the *crude* summary measure, since it removes the effect of the confounder.

Two other aspects of stratification should be noted. First, factors that are on the causal pathway between an exposure and a disease should not be regarded as confounding the association between the exposure and the outcome. Controlling for a factor that is on the causal pathway leads to underestimation of the strength of the effect of the exposure on the outcome under study. Occasionally, a variable that lies on the causal pathway may be adjusted for in the analysis if we want to assess whether the effect of the exposure on the outcome under study is entirely mediated through that intermediate variable or whether there may be other independent biological mechanisms. For instance, if we believe that human papillomavirus infection is on the causal pathway between number of sexual partners and cervical cancer, the association with number of sexual partners should disappear after adjusting for HPV infection. If the effect does not disappear completely, it would suggest that the effect of number of sexual partners on cervical cancer may be mediated through other biological mechanisms not directly related to HPV infection. In practice, this reasoning may not be so straightforward because of errors in the measurement of the intermediate variable.

Secondly, it is important to note that stratification assumes that each stratum is homogeneous in relation to the confounding variable. This assumption depends on both the validity of the data on the confounding variable (see Section 13.2) and the way strata are defined. For instance, if we control for age through stratification, this is better achieved if the strata are relatively narrow. Stratification into very broad age groups (e.g., 0–45 and 46+ years) is unlikely to be fully effective since, within each stratum, there are likely to be substantial differences in the age distribution of cases and controls (in a case–control study) or exposed and unexposed (in a cohort or intervention study).

## 14.4 Confounding and interaction

An underlying assumption in the calculation of a summary effect estimate is that the true effect is the same across strata and that any departures from this uniform effect are assumed to be due to random sampling variation. If there is substantial variation between the stratum-specific estimates of effect, this indicates the presence of *interaction* (also called *effect modification*) between the exposure of interest and the so-called confounding factor.

If there is interaction between the exposures under study and the confounder, a Mantel–Haenszel summary estimate of effect will be misleading, as it does not convey the full form of the exposure–outcome association, that is, that it varies according to the level of the stratifying variable. Thus, only if we are satisfied that the stratum-specific effect measures do not vary between themselves (i.e., there is no interaction), should we calculate an adjusted summary estimate of effect by taking a weighted mean of stratum-specific estimates. This concept is

***Example 14.4.*** *Suppose we are interested in examining the relationship between an exposure A and a certain outcome B in a cohort study. We start by calculating the crude rate ratio, which gives a value of 2.0. We then decide to examine the relationship between A and B separately for those who were exposed to a third variable C (stratum 1) and those who were not (stratum 2). Table 14.10 shows three possible results of this study and how they should be interpreted. In situation I, there is no confounding, because the crude and adjusted rate ratios are similar, and no interaction, because the rate ratios are similar for both strata. In situation II, there is confounding, because the crude and the adjusted rate ratios differ, but no interaction, because the effect is similar in the two strata. In situation III, there is strong interaction between A and C because the stratum-specific rate ratios are markedly different for those exposed and those not exposed to C.*

**Table 14.10.**
Example of confounding and interaction.

| | Crude rate ratio | Rate ratio in stratum 1 | Rate ratio in stratum 2 | Adjusted rate ratio | |
|---|---|---|---|---|---|
| Situation I | 2.0 | 2.0 | 2.0 | 2.0 | No confounding<br>No interaction |
| Situation II | 2.0 | 3.0 | 3.0 | 3.0 | Confounding present<br>No interaction |
| Situation III | 2.0 | 4.0 | 0.5 | – | Strong interaction |

not new. We discussed it briefly in Section 4.3.3 when we mentioned that age-standardized rates were an appropriate summary of the data only if the effect was homogeneous across the age-strata.

The first step in any stratified analysis is to assess whether interaction is present (Example 14.4). In most circumstances, this decision should be based on visual inspection of the data to examine the observed patterns of variation in the stratum-specific estimates. If they are the same, there is no interaction. In this situation, however, confounding may occur if the stratum-specific effects differ from the crude effect. If the stratum-specific estimates differ from each other, interaction is present and the emphasis in the analysis should be on describing how the association of interest is modified by the stratifying factor and all stratum-specific estimates of effect, and their confidence intervals, should be reported separately.

Deciding whether interaction exists or not in any particular analysis is often difficult, since the stratum-specific estimates are likely to vary because of random variation, even if the true effect is similar. For instance, the effect of Pap smear testing on cervical cancer in Example 14.3 was more marked for women of low educational level (RR = 0.20; 95% CI = 0.08–0.54) than for women of high educational level (RR = 0.40; 95% CI = 0.23–0.69), suggesting a weak interaction between Pap smear and educational level. However, this difference between the stratum-specific rate ratios may just reflect random variation. A variety of $\chi^2$ tests of heterogeneity are available to test the null hypothesis that the degree of variability in the series of stratum-specific estimates is consistent with random variation. In practice,

however, these tests are not very powerful. Epidemiological studies, unless they are specifically designed to do this, rarely have enough statistical power to detect interactions (see Section 15.4.4). Thus, more is usually gained by visual inspection of the size and pattern of the effect estimates across strata than from tests for interaction (effect modification).

## 14.5 Using the Mantel–Haenszel method to adjust for several confounders

In practice, it is frequently necessary to examine and possibly adjust for several confounders. This can be achieved by using the Mantel–Haenszel method, although, as the number of confounders increases, this method becomes impractical because most strata will have very sparse data. As we shall see later in this chapter, regression modelling techniques are more efficient methods in these circumstances.

***Example 14.5.*** *Assume that a case–control study was carried out to examine risk factors for human papillomavirus (HPV) infection of the cervix uteri. A standard questionnaire was used to collect detailed information on sociodemographic variables and sexual behaviour from 188 HPV-positive female cases and 571 HPV-negative female controls. Table 14.11 shows the distribution of cases and controls by smoking and HPV infection.*

**Table 14.11.**
Results from a case-control study on smoking and HPV infection: hypothetical data.

| | Smoking | | Total |
|---|---|---|---|
| | **Yes** | **No** | |
| HPV-positive | 42 (*a*) | 146 (*b*) | 188 ($n_1$) |
| HPV-negative | 90 (*c*) | 481 (*d*) | 571 ($n_0$) |
| **Total** | **132 ($m_1$)** | **627 ($m_0$)** | **759 (*N*)** |
| Crude odds ratio = (42/146) / (90/481) = 1.54 | | | |
| 95% confidence interval = 1.02–2.32 | | | |
| $\chi^2$ = 4.25 on 1d.f.; *P* = 0.039 | | | |

In Example 14.5, the *crude odds ratio* is 1.54. The 95% confidence interval and the *P*-value suggest that chance is an unlikely explanation of this finding. Thus, women who smoked were more likely to be HPV-positive than those who did not.

Number of sexual partners is a well known risk factor for HPV infection and it may have confounded the association between smoking and HPV infection. Table 14.12 shows the association found between smoking and HPV stratified by reported number of lifetime sexual partners (categorized as < 2 partners and ≥ 2 partners).

Examination of the stratum-specific odds ratios (and their confidence intervals) suggests that the effect of smoking on HPV infection in women who reported less than two sexual partners is similar to the

**< 2 sexual partners**

| | Smoking | | Total |
|---|---|---|---|
| | **Yes** | **No** | |
| HPV-positive | 31 ($a_1$) | 124 ($b_1$) | 155 ($n_{11}$) |
| HPV-negative | 78 ($c_1$) | 437 ($d_1$) | 515 ($n_{01}$) |
| **Total** | **109 ($m_{11}$)** | **561 ($m_{01}$)** | **670 ($N_1$)** |

Odds ratio = 1.40

95% confidence interval = 0.88–2.22

**≥ 2 sexual partners**

| | Smoking | | Total |
|---|---|---|---|
| | **Yes** | **No** | |
| HPV-positive | 11 ($a_2$) | 22 ($b_2$) | 33 ($n_{12}$) |
| HPV-negative | 12 ($c_2$) | 44 ($d_2$) | 56 ($n_{02}$) |
| **Total** | **23 ($m_{12}$)** | **66 ($m_{02}$)** | **89 ($N_2$)** |

Odds ratio = 1.83

95% confidence interval = 0.70–4.80

$\chi^2$ test for heterogeneity = 0.24 on 1 d.f.; $P = 0.63$

$$OR_{MH} = \frac{\Sigma a_i d_i / N_i}{\Sigma b_i c_i / N_i} = \frac{(31\times437)/670 + (11\times44)/89}{(124\times78)/670 + (22\times12)/89} = \frac{25.66}{17.40} = 1.47$$

95% confidence interval = 0.95–2.26

$\chi^2 = 2.97$ on 1d.f.; $P = 0.09$

**Table 14.12.**
Hypothetical case–control study on smoking and HPV infection described in Example 14.5: results stratified by number of lifetime sexual partners.

effect in those who reported two or more. This is confirmed by the $\chi^2$ test for heterogeneity. Since the effect of smoking on HPV infection is uniform across the two strata, it is appropriate to calculate a pooled adjusted odds ratio. Thus, after adjusting for number of sexual partners, the odds ratio of smoking is reduced a little from 1.54 to 1.47. This result suggests that this variable was a weak confounder of the association between smoking and HPV infection. But even after adjusting for number of sexual partners, the 95% confidence interval is still consistent with an effect of smoking on HPV infection. Thus, these results suggest that smoking increased the risk of HPV infection, but this risk was marginally weaker after allowing for the confounding effect of number of sexual partners.

The same technique can be used to obtain an odds ratio adjusted for age. In this hypothetical study, age (in years) was categorized into six groups (< 20, 20–24, 25–29, 30–34, 35–44, ≥ 45). Thus, we need to construct six 2 × 2 tables of smoking and HPV infection, one for each age-group. The cells of these 2 × 2 tables are shown in Table 14.13.

The $OR_{MH}$ adjusted for age is slightly lower than the crude odds ratio (1.47 versus 1.54), suggesting that age was also a weak confounder of the association between smoking and HPV infection.

**Table 14.13.**
Hypothetical case–control study on smoking and HPV infection described in Example 14.5: results stratified by age-group.

| Stratum | Age (years) | $a_i$ | $b_i$ | $c_i$ | $d_i$ | $N_i$ | $a_i d_i/N_i$ | $b_i c_i/N_i$ |
|---|---|---|---|---|---|---|---|---|
| 1 | <20 | 3 | 10 | 16 | 79 | 108 | 2.19 | 1.48 |
| 2 | 20–24 | 13 | 44 | 16 | 92 | 165 | 7.25 | 4.27 |
| 3 | 25–29 | 6 | 33 | 22 | 62 | 123 | 3.02 | 5.90 |
| 4 | 30–34 | 8 | 25 | 10 | 75 | 118 | 5.08 | 2.12 |
| 5 | 35–44 | 8 | 22 | 18 | 89 | 137 | 5.20 | 2.89 |
| 6 | ≥45 | 4 | 12 | 8 | 84 | 108 | 3.11 | 0.89 |
| All strata | | 42 | 146 | 90 | 481 | 759 | 25.85 | 17.55 |

$\chi^2$ test for heterogeneity = 7.37 on 5 d.f.; $P = 0.20$

$$OR_{MH} = \frac{\Sigma a_i d_i/N_i}{\Sigma b_i c_i/N_i} = \frac{25.85}{17.55} = 1.47$$

95% confidence interval = 0.96–2.32
$\chi^2 = 3.06$ on 1d.f.; $P = 0.08$

Thus, so far, we have estimated the effect of smoking on HPV infection adjusted for number of sexual partners and the effect of smoking adjusted for age. We could estimate the effect of smoking on HPV infection adjusted *simultaneously* for number of partners and age using the Mantel–Haenzsel method. To do this, we need to construct a 2 × 2 table of smoking by HPV infection for every possible combination of number of partners and age-group. Since number of partners forms two strata (<2 and ≥2) and age forms six strata (< 20, 20–24, 25–29, 30–34, 35–44, ≥ 45), we need to construct 2 × 6 =12 such 2 × 2 tables. The cells of the 2 × 2 tables for smoking for these 12 strata are shown in Table 14.14. Thus, after adjusting for the confounding effects of number of partners and age, the effect of smoking on HPV infection is even smaller ($OR_{MH} = 1.41$).

The Mantel–Haenszel method can be extended to adjust simultaneously for more than two confounders. For example, to estimate the effect of smoking on HPV infection, allowing for number of sexual partners (two strata), age (six strata), marital status (three strata: married, single, widowed/divorced) and educational level (three strata), we would construct 2×6×3×3=108 2×2 tables. Clearly, in 108 tables formed from a data-set of 188 cases and 571 controls, some strata will have very small numbers of observations, if any.

A further problem with the Mantel–Haenszel method is that each explanatory variable included in the analysis has to be classified as either an exposure or a confounder, and there may be only one exposure. For example, smoking was our exposure, and number of partners and age were our confounders. We therefore obtained an odds ratio for smoking adjusted for number of partners and age. These results did not give us the odds ratio for number of partners adjusted for smoking and age, or the odds ratios for age adjusted for smoking and number of partners. These would have required further Mantel–Haenszel analyses.

| Stratum | Age (yrs) | Number of partners | $a_i$ | $b_i$ | $c_i$ | $d_i$ | $N_i$ | $a_i d_i/N_i$ | $b_i c_i/N_i$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | <20 | <2 | 2 | 8 | 11 | 67 | 88 | 1.52 | 1.00 |
| 2 | 20–24 | <2 | 10 | 36 | 14 | 85 | 145 | 5.86 | 3.48 |
| 3 | 25–29 | <2 | 3 | 28 | 19 | 59 | 109 | 1.62 | 4.88 |
| 4 | 30–34 | <2 | 5 | 21 | 10 | 65 | 101 | 3.22 | 2.08 |
| 5 | 35–44 | <2 | 7 | 19 | 17 | 78 | 121 | 4.51 | 2.67 |
| 6 | ≥45 | <2 | 4 | 12 | 7 | 83 | 106 | 3.13 | 0.79 |
| 7 | <20 | ≥2 | 1 | 2 | 5 | 12 | 20 | 0.60 | 0.50 |
| 8 | 20–24 | ≥2 | 3 | 8 | 2 | 7 | 20 | 1.05 | 0.80 |
| 9 | 25–29 | ≥2 | 3 | 5 | 3 | 3 | 14 | 0.64 | 1.07 |
| 10 | 30–34 | ≥2 | 3 | 4 | 0 | 10 | 17 | 1.76 | 0.00 |
| 11 | 35–44 | ≥2 | 1 | 3 | 1 | 11 | 16 | 0.69 | 0.19 |
| 12 | ≥45 | ≥2 | 0 | 0 | 1 | 1 | 2 | 0.00 | 0.00 |
| All strata | | | 42 | 146 | 90 | 481 | 759 | 24.60 | 17.46 |

$\chi^2$ test for heterogeneity = 12.44 on 10 d.f.; $P = 0.26$

$$OR_{MH} = \frac{\Sigma a_i d_i/N_i}{\Sigma b_i c_i/N_i} = \frac{24.60}{17.46} = 1.41$$

95% confidence interval = 0.91–2.24

$\chi^2 = 2.62$ on 1 d.f.; $P = 0.132$

**Table 14.14.** Hypothetical case–control study on smoking and HPV infection described in Example 14.5: results stratified by age and lifetime number of sexual partners.

## 14.6 Using regression modelling to adjust for the effect of confounders

Regression models summarize the relationship between an outcome (also called dependent) variable and several explanatory (independent) variables as a mathematical equation. The general form of this equation is

Outcome variable = function (explanatory variables)

There are many types of regression model. The choice of any particular model depends on the characteristics of the outcome variable (i.e., continuous or categorical) and on the way it is mathematically related to the explanatory variables. The simplest mathematical model we could use has already been introduced in Section 11.2.1 to describe the relationship between two quantitative variables.

A discussion of these models and the assumptions underlying them is beyond the scope of this chapter. However, these modelling techniques are commonly used in epidemiological studies and, therefore, in the rest of this chapter we will try to illustrate how these techniques relate to the Mantel–Haenszel method, to allow the reader to understand and interpret results from published work where they have been used.

Let us consider again the hypothetical case–control study described in Example 14.5. We can use a particular regression technique, called *logistic regression*, to analyse data from unmatched case–control studies. In this analysis, we start by using a logistic regression model which includes as explanatory variables only the exposure under study–smoking. The results are shown in Table 14.15 (model 1). The odds ratio estimated by this logistic regression model is the same as the crude odds ratio we obtained in Table 14.11.

**Table 14.15.**
Hypothetical case–control study on risk factors for HPV cervical infection described in Example 14.5. Results obtained from logistic regression models which included an increasing number of explanatory variables: smoking, lifetime number of sexual partners and age. (The values underlined correspond to those obtained earlier with the Mantel–Haenszel technique shown in Table 14.11, 14.12 and 14.14.)

| Variable | Odds ratio | 95% confidence interval |
|---|---|---|
| *Model 1* | | |
| smoking[a] | 1.54 | 1.02–2.32 |
| *Model 2* | | |
| smoking[a] | 1.47 | 0.97–2.23 |
| partners[b] | 1.90 | 1.19–3.03 |
| *Model 3* | | |
| smoking[a] | 1.43 | 0.93–2.19 |
| partners[b] | 1.95 | 1.19–3.17 |
| age 2[c] | 4.14 | 2.12–8.12 |
| age 3 | 3.58 | 1.77–7.24 |
| age 4 | 3.01 | 1.47–6.15 |
| age 5 | 2.18 | 1.07–4.46 |
| age 6 | 1.49 | 0.67–3.33 |

[a] Categorized as 'non-smokers' (baseline) and 'smokers'.

[b] Categorized as '< 2 partners' (baseline) and '≥ 2 partners'

[c] Categorized as age1 = <20 (baseline); age2 = 20–24; age3 = 25–29; age4 = 30–34; age5 = 35–44; age6 = 45+ years.

We then move on to use a model which includes both smoking and number of sexual partners as explanatory variables (model 2 in Table 14.15). This model gives

> Odds ratio for smoking versus non-smoking adjusted for number of sexual partners = 1.47 (95% CI = 0.97–2.23)

Thus, after adjusting for number of sexual partners, the effect of smoking on HPV infection became smaller (1.47 versus 1.54). This result is similar to that obtained earlier when we used the Mantel–Haenszel technique to control for the effect of number of sexual partners (Table 14.12). But in contrast to this technique, this regression model also gives us the odds ratio for number of sexual partners adjusted for smoking:

> Odds ratio for ≥2 sexual partners versus <2, adjusted for smoking = 1.90 (95% CI = 1.19–3.03)

Thus, there is a statistically significant increased risk of HPV infection associated with two or more sexual partners, even after taking differences in smoking into account.

We can use a more complex model which includes smoking, number of sexual partners and age (model 3, Table 14.15). This model gives us an estimate of the effect of smoking on HPV infection adjusted for number of sexual partners and age of 1.43 (95% CI = 0.93–2.19), which is similar to that obtained before with the Mantel–Haenszel method ($OR_{MH}$ = 1.41; 95% CI = 0.91–2.24) (see Table 14.14). However, unlike the Mantel–Haenszel analysis, the logistic regression analysis also gives the following extra odds ratios:

Odds ratio for ≥ 2 sexual partners versus < 2, adjusted for smoking and age = 1.95 (95% CI = 1.19–3.17)

Odds ratio for age 20–24 versus age <20, adjusted for smoking and partners = 4.14 (95% CI = 2.12–8.12)

Odds ratio for age 25–29 versus age <20, adjusted for smoking and partners = 3.58 (95% CI = 1.77–7.24)

Odds ratio for age 30–34 versus age <20, adjusted for smoking and partners = 3.01 (95% CI = 1.47–6.15)

Odds ratio for age 35–44 versus age <20, adjusted for smoking and partners = 2.18 (95% CI = 1.07–4.46)

Odds ratio for age 45+ versus age <20, adjusted for smoking and partners = 1.49 (95% CI = 0.67–3.33)

One of the main advantages of regression modelling is that it does not require us to define which explanatory variable is the exposure and which ones are the potential confounders, since all explanatory variables are treated in the same way. This is particularly important in studies designed to examine the effect of a wide range of exposures rather than just the effect of a specific one.

Similar regression models can be applied to data from studies of other designs. Let us consider again the hypothetical cohort study on Pap smear use and cervical cancer described in Example 14.3 (Section 14.2.3). In Table 14.8, we calculated the crude rate ratio to measure the effect of Pap smear testing on cervical cancer. We then went on to calculate the effect of Pap smear adjusting for educational level (Table 14.9) using the Mantel–Haenszel technique. We can also use the Mantel–Haenszel technique to adjust simultaneously for educational level, marital status and age at first intercourse. The results are shown in Table 14.16.

Using the Mantel–Haenszel technique to adjust simultaneously for educational level, marital status and age at first intercourse involved the formation of 18 (=2×3×3) strata. However, only seventeen cervical cancer cases occurred in women who reported having ever had a Pap smear. Consequently, there were empty cells in several strata.

**Table 14.16.**
Hypothetical cohort study on Pap smear testing and cervical cancer described in Example 14.3. Results obtained using the Mantel–Haenszel technique.

| Variables adjusted for | Cervical cancer rate ratio for Pap smear use | 95% confidence interval |
|---|---|---|
| None | 0.32 (crude) | 0.20–0.52 |
| Educational level[a] | 0.32 | 0.20–0.52 |
| Educational level and marital status[b] | 0.40 | 0.25–0.66 |
| Educational level, marital status and age at first intercourse[c] | 0.43 | 0.27–0.72 |

[a] Categorized as 'low educational level' and 'high educational level'.

[b] Categorized as marital status 1=married; 2=single; 3=divorced/widowed.

[c] Categorized as age at first intercourse 1=<18 years; 2=18–22 years; 3=22+ years.

We can use a special regression modelling technique, called *Poisson regression*, to analyse the data from this hypothetical cohort study on Pap smear and cervical cancer. The results are shown in Table 14.17.

As with logistic regression, we start by using a model which includes only one explanatory variable, Pap smear use (model 1 in Table 14.17). This model gives us the Poisson estimate of the cervical cancer rate ratio for Pap smear use. This value corresponds to the crude cervical cancer rate ratio obtained earlier. This model gives us another value called 'constant', which corresponds to the cervical cancer incidence rate in women who reported never having had a Pap smear (see Table 14.8). From these values, we can calculate the incidence rate in the exposed as $0.00074 \times 0.32 = 0.0002368 = 24$ per 100 000 person-years (the same as the value obtained in Table 14.8).

We then move on to add to our model another explanatory variable, for instance, educational level (model 2 in Table 14.17). This model gives us the Poisson estimate of the cervical cancer rate ratio for Pap smear use adjusted for educational level, which is 0.32 (95% CI = 0.20–0.52). This is the same value obtained when the Mantel–Haenszel technique was used to control for educational level (Table 14.16). In contrast with the Mantel–Haenszel technique, this model also gives us

> Cervical cancer rate ratio for high versus low educational level adjusted for Pap smear use = 0.73 (95% CI = 0.65–0.82)

In the last model (model 4) shown in Table 14.17, we included Pap smear use, educational level, marital status and age at first intercourse as explanatory variables. The Poisson estimate of the rate ratio for Pap smear use adjusted for educational level, marital status and age at first intercourse is 0.46 (95% CI = 0.29–0.75), similar to that obtained with the Mantel–Haenszel method ($RR_{MH} = 0.43$; 95% CI = 0.27–0.72) (Table 14.16). But with the Poisson regression, we also obtained the following additional rate ratios:

> Cervical cancer rate ratio for high versus low educational level adjusted for Pap smear use, marital status and age at first intercourse = 0.77 (95% CI = 0.68–0.87)

| Variable | Baseline rate (per person year) | Cervical cancer rate ratio | 95% confidence interval |
|---|---|---|---|
| *Model 1* | | | |
| constant | 0.00074 | | 0.0007–0.0008 |
| Pap smear use[a] | | 0.32 | 0.20–0.52 |
| *Model 2* | | | |
| constant | 0.008 | | 0.008–0.009 |
| Pap smear use[a] | | 0.32 | 0.20–0.52 |
| Educational level[b] | | 0.73 | 0.65–0.82 |
| *Model 3* | | | |
| constant | 0.005 | | 0.004–0.005 |
| Pap smear use[a] | | 0.41 | 0.25–0.66 |
| Educational level[b] | | 0.74 | 0.66–0.84 |
| Marital status2[c] | | 2.68 | 2.28–3.15 |
| Marital status3 | | 1.89 | 1.61–2.21 |
| *Model 4* | | | |
| constant | 0.008 | | 0.006–0.009 |
| Pap smear use[a] | | 0.46 | 0.29–0.75 |
| Educational level[b] | | 0.77 | 0.68–0.87 |
| Marital status2[c] | | 2.68 | 2.27–3.15 |
| Marital status3 | | 1.60 | 1.36–1.87 |
| Age at first intercourse2[d] | | 0.52 | 0.46–0.59 |
| Age at first intercourse3 | | 0.13 | 0.09–0.19 |

[a] Categorized as 'never' (baseline) and 'ever'.
[b] Categorized as 'low educational level' (baseline) and 'high educational level'.
[c] Categorized as marital status 1=married (baseline), 2=single, 3=divorced/widowed.
[d] Categorized as age at first intercourse 1= < 18 years (baseline), 2=18–22 years, 3=22+ years.

**Table 14.17.**
Hypothetical cohort study on Pap smear and cervical cancer described in Example 14.3. Results obtained from Poisson regression models with increasing numbers of explanatory variables. (The values underlined correspond to those obtained with the Mantel–Haenszel technique shown in Table 14.16.)

Cervical cancer rate ratio for single versus married women adjusted for Pap smear use, educational level and age at first intercourse = 2.68 (95% CI = 2.27–3.15)

Cervical cancer rate ratio for divorced/widowed versus married women adjusted for Pap smear use, educational level and age at first intercourse = 1.60 (95% CI = 1.36–1.87)

Cervical cancer rate ratio for age at first intercourse 18–22 versus < 18 years adjusted for Pap smear use, educational level and marital status = 0.52 (95% CI = 0.46–0.59)

Cervical cancer rate ratio for age at first intercourse 22+ versus < 18 years adjusted for Pap smear use, educational level and marital status = 0.13 (95% CI = 0.09–0.19)

*Logistic regression* is used for estimating odds ratios. It may therefore be used in a case–control study, a cross-sectional study, or if estimating 'risks' rather than 'rates', in a cohort study. *Poisson regression* models are used for estimating rate ratios using person-time data. Other commonly used regression models in epidemiology are:

*Conditional logistic regression*: logistic regression analysis is suitable for unmatched case–control studies or frequency-matched case–control studies. Individually matched case–control studies require a slightly different approach called conditional logistic regression analysis. This modelling technique is the only way we can adjust for confounders other than the matching factor(s) used in the design of these studies.

*Cox's proportional hazards model*: this type of regression model is used when the *time to an event* is of particular interest (as in survival analysis).

## 14.7 Conclusions

In summary, the Mantel–Haenszel method is a very useful technique to adjust for confounders, and this approach is often adequate for data with few confounders. However, in order to adjust simultaneously for several confounders, regression modelling methods may be necessary.

It is important, however, to stress that any analysis should start by using the Mantel–Haenszel method to obtain preliminary crude effect estimates and effect estimates adjusted for each confounder separately. The cross-tabulations used for stratification in this technique allow the investigator to observe most of the important relationships and interactions that are present and to detect errors and inconsistencies in the data that might not otherwise be evident.

Regression models can then be used in a second stage of the analysis to adjust simultaneously for several confounders. One of the main disadvantages of regression modelling is that we lose sight of the data, so that it is often regarded as a 'black box' approach. Statistical modelling should not be used by people who are not familiar with it and who do not understand the assumptions upon which it is based.

## Box 14.1. Key issues

- Any analysis of data should be planned carefully. In general, it should involve the following steps:

    1. Produce simple tables to check the consistency of the data.
    2. Calculate crude measures of effect.
    3. Stratify by levels of the potential confounding factor.
    4. Compute stratum-specific effect estimates.
    5. Check uniformity of the stratum-specific estimates by visual inspection or by performing tests of statistical significance.
    6. If the effect is thought to be uniform across strata, calculate a pooled adjusted summary estimate of the effect using the Mantel–Hanszel method. Calculate confidence intervals for the adjusted estimate and the Mantel–Hanszel $\chi^2$ test.
    7. If the effect is not thought to be uniform (i.e., if interaction is present), report stratum-specific estimates, confidence intervals and $\chi^2$ for each estimate.
    8. Use regression modelling techniques to adjust simultaneously for several confounders.

- The simple classical methods based on stratification should always be used in the initial phase of an analysis. The cross-tabulations used in stratification keep the investigator in touch with the data.

- Regression models can be used in a second stage of the analysis to adjust simultaneously for several confounders. In contrast to the classical methods, regression modelling is, to a certain extent, a 'black box' approach and because of this, it may lead to serious errors. These methods are complex statistical procedures that should never be used by those who are unfamiliar with them.

## Further reading

* In this chapter, we have presented formulae to calculate Mantel–Haenszel rate ratios, risk ratios and odds ratios. Formulae to calculate adjusted estimates of risk and rate differences can be found in Greenland & Robins (1985).

* Stratification and regression modelling techniques are covered in a much more comprehensive (although more statistically elaborate) way in Breslow & Day (1980, 1987) and Clayton & Hills (1993).

# Appendix 14.1.
# Confidence intervals and statistical tests for adjusted relative measures of effect

Note that there are several methods to calculate confidence intervals for adjusted relative measures of effect, which may yield slightly different values from those obtained here. These calculations can easily be performed by statistical computing packages such as EPI INFO, STATA or EGRET.

## A14.1.1 Adjusted odds ratio

### *Confidence interval for Mantel–Haenszel odds ratio*

The standard error (SE) of the logarithm of the Mantel–Haenszel odds ratio ($OR_{MH}$) can be estimated as

$$SE\ (\ln OR_{MH}) = \sqrt{\frac{\Sigma(b_i c_i/N_i)^2\, v_i}{(\Sigma b_i c_i/N_i)^2}}$$

in which $\Sigma b_i c_i/N_i$ corresponds to the denominator of the formula used to calculate the $OR_{MH}$ and $v_i = 1/a_i + 1/b_i + 1/c_i + 1/d_i$

Thus, in the ovarian cancer study (Example 14.1; Table 14.2), we have

$$\Sigma b_i c_i/N_i = (32 \times 8)/77 + (4 \times 50)/81 = 3.32 + 2.47 = 5.79$$
$$v_1 = 1/9 + 1/32 + 1/8 + 1/28 = 0.30$$
$$v_2 = 1/15 + 1/4 + 1/50 + 1/12 = 0.42$$
$$\Sigma(b_i c_i/N_i)^2\, v_i = (3.32^2 \times 0.30) + (2.47^2 \times 0.42) = 3.31 + 2.56 = 5.87$$

Thus,

$$SE\ (\ln OR_{MH}) = \sqrt{5.87/(5.79)^2} = 0.42$$

An 'approximate' 95% confidence interval for the $\ln OR_{MH}$ can then be estimated as

$$95\%\ CI\ (\ln OR_{MH}) = (\ln 0.95) \pm 1.96 \times 0.42 = -0.05 \pm 1.96 \times 0.42 = -0.87 \text{ to } 0.77$$

An 'approximate' 95% confidence interval for the $OR_{MH}$ can be obtained by taking anti-logarithms:

$$95\%\ CI\ (OR_{MH}) = e^{-0.87} \text{ to } e^{0.77} = 0.42 \text{ to } 2.16$$

### *The Mantel–Haenszel $\chi^2$ test*

The Mantel–Haenszel $\chi^2$ test can be used to determine whether $OR_{MH}$ is significantly different from one. The test is just an extension of the Mantel–Haenszel $\chi^2$ test for a single $2 \times 2$ table presented in Section A6.1.3. The null hypothesis is that there is no association between the exposure and the disease (that is, the odds ratio is equal to one) within any of the individual strata. In order to perform this test, we must first obtain the following, from each stratum $i$:

(i) The observed value of $a_i$:

$$O(a_i) = a_i$$

(ii) The expected value of $a_i$, assuming the null hypothesis of no association:

$$E(a_i) = n_{1i}m_{1i}/N_i$$

(iii) The variance of $a_i$, assuming the null hypothesis of no association:

$$V(a_i)=n_{1i}n_{0i}m_{1i}m_{0i}/(N_i^2(N_i - 1))$$

We then sum each of these quantities over all the strata. In Example 14.1 (Table 14.2), we obtain

$$\sum O(a_i) = 9 + 15 = 24$$

$$\sum E(a_i) = (41 \times 17)/77 + (19 \times 65)/81\ 9.05 + 15.25 = 24.30$$

$$\sum V(a_i)=(41\times36\times17\times60)/(77^2\times76) + (19\times62\times65\times16)/(81^2\times80) = 3.34 + 2.33 = 5.67$$

We would expect the difference between our observed and expected values to be small if the null hypothesis were true. To test whether the differences obtained are greater than would be expected by chance, we calculate

$$\chi^2 = (\sum O(a_i) - \sum E(a_i))^2 / \sum V(a_i)$$

and obtain a $P$-value by referring our result to the $\chi^2$ distribution with one degree of freedom (d.f.).

In our example, $\chi^2 = (24 - 24.30)^2/5.67 = 0.016$ on 1 d.f. This gives $P = 0.93$, from which we conclude that after adjusting for oral contraceptive use, there is no evidence of any association between smoking and ovarian cancer. So $OR_{MH} = 0.95$ is not statistically significantly different from 1.

## A14.1.2 Adjusted risk ratio

### *Confidence interval for Mantel–Haenszel risk ratio*

The standard error (SE) of the logarithm of Mantel–Haenszel risk ratio ($R_{MH}$) can be estimated by

$$\text{SE}(\ln R_{MH}) = \sqrt{\frac{\Sigma(n_{1i}m_{1i}m_{0i} - a_i b_i N_i)/N_i^2}{(\Sigma a_i m_{0i}/N_i)(\Sigma b_i m_{1i}/N_i)}}$$

where the expressions $\sum a_i m_{0i}/N_i$ and $\sum b_i m_{1i}/N_i$ correspond, respectively, to the numerator and denominator of the formula used to calculate $R_{MH}$.

In our occupational study (Example 14.2; Table 14.6)

$$\Sigma a_i m_{0i}/N_i = 264$$

$$\Sigma b_i m_{1i}/N_i = 132$$

$$\Sigma(n_{1i}m_{1i}m_{0i} - a_i b_i N_i)/N_i^2 = (240\times4000\times16\,000 - 80\times160\times20\,000)/20\,000^2 + (600\times80\,000\times80\,000 - 400\times200\times160\,000)/160\,000^2$$

$$= 37.76 + 149.50 = 187.26$$

Thus,

$$\text{SE}(\ln R_{MH}) = \sqrt{187.26/(264 \times 132)} = 0.073$$

$$95\%\text{ confidence interval } (\ln R_{MH}) = (\ln 2.0) \pm 1.96 \times 0.073 = 0.547 \text{ to } 0.833$$

An 'approximate' 95% confidence interval for $R_{MH}$ can be obtained by taking anti-logarithms:

$$95\%\text{ confidence interval } (R_{MH}) = e^{0.547} \text{ to } e^{0.833} = 1.73 \text{ to } 2.30$$

### *The Mantel–Haenszel $\chi^2$ test*

The Mantel–Haenszel $\chi^2$ test to assess whether $R_{MH}$ is statistically significantly different from unity is similar to that used above for odds ratio. The null hypothesis is that there is no association between the exposure and the disease within any of the individual strata, that is that the risk ratio is equal to one in each stratum.

In our occupational study (Example 14.2; Table 14.6)

$$\Sigma O(a_i) = 80 + 400 = 480$$

$$\Sigma E(a_i) = \Sigma n_{1i}m_{1i}/N_i = 240\times4000/20\,000 + 600\times80\,000/160\,000 = 48 + 300 = 348$$

$$\Sigma V(a_i) = \Sigma n_{1i}n_{0i}m_{1i}m_{0i}/(N_i^2(N_i-1)) = (240\times19\,760\times4000\times16\,000)/(20\,000^2\times19\,999) + (600\times159\,400\times80\,000\times80\,000)/(160\,000^2\times159\,999)$$

$$= 37.94 + 149.44 = 187.38$$

We can now calculate the $\chi^2$ test:

$$\chi^2 = (\Sigma O(a_i) - \Sigma E(a_i))^2 / \Sigma V(a_i) = (480 - 348)^2/187.38 = 92.99$$

and obtain a *P*-value by referring our result to the $\chi^2$ distribution with 1 d.f. In this example, $P < 0.0001$. Thus, it is very unlikely that these results are due to chance.

## A14.1.3 Adjusted rate ratio

### *Confidence interval for Mantel–Haenszel rate ratio*

As with any other estimate, it is useful to be able to construct a 95% confidence interval round a Mantel–Haenszel rate ratio ($RR_{MH}$). The standard error of the logarithm of a rate ratio can be estimated by

$$SE\ (\ln RR_{MH}) = \sqrt{\frac{\Sigma V(a_i)}{(\Sigma a_i y_{0i}/y_i)(\Sigma b_i y_{1i}/y_i)}}$$

We can obtain an 'approximate' 95% confidence interval around the logarithm of the $RR_{MH}$ as

$$\ln RR_{MH} \pm 1.96 \times SE\ (\ln RR_{MH})$$

An 'approximate' 95% confidence interval for $RR_{MH}$ can then be obtained by taking anti-logarithms.

Note that $\Sigma a_i y_{0i}/y_i$ and $\Sigma b_i y_{1i}/y_i$ are, respectively, the numerator and the denominator of the formula for the Mantel–Haenszel rate ratio, which were calculated for Example 14.3 in Section 14.2.3. Thus, we only need to calculate $\Sigma V(a_i)$:

$$\Sigma a_i y_{0i}/y_i = 16.24$$

$$\Sigma b_i y_{1i}/y_i = 50.23$$

$$V(a_i) = \Sigma n_i y_{1i} y_{0i}/y_i^2 = (710 \times 38\ 346 \times 828\ 149)/866\ 495^2 + (431 \times 32\ 838 \times 690\ 552)/723\ 390^2 = 30.03 + 18.68 = 48.71$$

$$\ln RR_{MH} = \ln (0.32) = -1.14$$

$$SE\ (\ln RR_{MH}) = \sqrt{48.71/(16.24 \times 50.23)} = 0.244$$

$$95\%\ \text{confidence interval}\ (\ln RR_{MH}) = -1.14 \pm 1.96 \times 0.244 = -1.62\ \text{to}\ -0.66$$

$$95\%\ \text{confidence interval}\ (RR_{MH}) = e^{-1.62}\ \text{to}\ e^{-0.66} = 0.20\ \text{to}\ 0.52$$

### *The Mantel–Haenszel $\chi^2$ test*

We may want to perform a significance test to test whether the true rate ratio is different from one. Our null hypothesis is that the true rate ratio in *all* strata is one. In order to calculate the test, we need first to compute the following for each stratum *i*:

(i) Observed value of $a_i = O(a_i) = a_i$

(ii) Expected value of $a_i = E(a_i) = n_i y_{1i}/y_i$

(iii) Variance of $a_i = V(a_i) = n_i y_{1i} y_{0i}/y_i^2$

An overall test of significance (that the common rate ratio is unity) is given by

$$\chi^2 = (\Sigma O(a_i) - \Sigma E(a_i))^2 / \Sigma V(a_i)$$

where the summation is over all strata. The value calculated should be looked up in tables of the $\chi^2$ distribution with one degree of freedom.

Thus, in Example 14.3,

$$\Sigma O(a_i) = 13 + 4 = 17$$

$$\Sigma E(a_i) = (710 \times 38\,346/866\,495) + (431 \times 32\,838/723\,390)$$
$$= 31.42 + 19.57 = 50.99$$

$\Sigma V(a_i) = 48.71$, which was obtained above for the calculation of the confidence interval.

$$\chi^2 = (17 - 50.99)^2/48.71 = 23.72$$

This $\chi^2$ on 1 d.f. corresponds to $P < 0.001$.