18
Comparison of odds within strata

This chapter deals with methods for analysing stratified case-control studies which closely parallel the methods for cohort studies discussed in Chapter 15.

18.1 The constant odds ratio model

As an example we return to the study of the effect of BCG vaccination upon the incidence of leprosy. Since leprosy incidence increases with age among young people, age is certainly a variable which would have been controlled in an experiment. In Chapter 16 it was shown that BCG-vaccinated individuals had just under one half of the incidence of leprosy as compared with unvaccinated persons, but age was ignored in the analysis. This could have biased the estimated effect of BCG vaccination because BCG vaccination in the area (Northern Malawi) was introduced gradually in infants and young children, so that people who were older during the study period, having been born at earlier dates, were less likely to have been vaccinated. As a result, on average the vaccinated group will be younger than the unvaccinated group. This means that, even if BCG vaccination were totally ineffective, one would expect to observe lower rates in vaccinated members of the base cohort, simply as a result of their relative youth.

Table 18.1 subdivides these data by strata corresponding to 5-year age

Table 18.1. BCG vaccination and leprosy by age

	BCG scar				Odds
	Leprosy cases		Healthy population		ratio
Age	Absent	Present	Absent	Present	estimate
0–4	1	1	7593	11719	0.65
5–9	11	14	7143	10184	0.89
10–14	28	22	5611	7561	0.58
15–19	16	28	2208	8117	0.48
20–24	20	19	2438	5588	0.41
25–29	36	11	4356	1625	0.82
30–34	47	6	5245	1234	0.54

bands. The table also shows age-specific odds ratios. Although there is random variation, there is no systematic trend of the odds ratio with age, and it seems reasonable to make the assumption that the odds ratio parameter is the same in all age bands. In the next section we show how an estimate of this common odds ratio can be calculated.

18.2 An estimate of the common odds ratio

In the prospective approach to the analysis, the assumption of a common odds ratio implies that ω_1^t/ω_0^t is constant, so that the model can be expressed in terms of the odds ratio parameter θ and the ω_0^t parameters. Alternatively, in the retrospective approach the model is expressed in terms of θ and the parameters Ω_0^t. In both approaches, replacing the nuisance parameters by their estimates leads to the profile likelihood for θ. If there are not too many strata, and the data are not too sparse in each stratum, then the profile likelihood for θ can be used to find the most likely value and the supported range. For coarsely stratified data sets such as Table 18.1, these conditions are met. Such an analysis is not feasible by hand, but would usually be carried out on a computer using *logistic regression* (see Chapter 23).

When the data are very finely stratified so that each stratum contains very few cases and controls, the profile likelihood approach can be unreliable, and the hypergeometric likelihood should be used. The total log likelihood is then obtained by adding together the hypergeometric log likelihoods for the different strata. Again, the most likely value M and the standard deviation S cannot usually be computed by hand, but would be carried out using a *conditional* logistic regression program (see Chapter 29). However, the calculations for the score test for $\theta = 1$ are straightforward. For a single stratum the score under the hypergeometric likelihood is

$$U = D_1 - E_1$$

where D_1 is the observed number of exposed cases and $E_1 = DN_1/N$ is the expected number under the null hypothesis. The score variance is

$$V = \frac{DHN_0N_1}{(N)^2(N-1)}.$$

Since every stratum contributes additively to the overall log likelihood, the overall score is a sum of contributions from each stratum of exactly the same form as above. Thus, the score is

$$U = \sum (D_1^t - E_1^t)$$

where

$$E_1^t = D^t \frac{N_1^t}{N^t}.$$

and the overall score variance is

$$V = \sum \frac{D^t H^t N_0^t N_1^t}{(N^t)^2(N^t-1)}.$$

Exercise 18.1. Show that the first age band in Table 18.1 makes a contribution of -0.21 to U and 0.48 to V.

The overall test statistic is obtained by repeating these calculations for each stratum and yields

$$U = -0.21 - 0.69 - 6.68 - 6.56 - 8.11 - 1.76 - 4.06 = -28.07$$

and

$$V = 0.48 + 6.05 + 12.18 + 7.38 + 8.22 + 9.22 + 8.09 = 51.62.$$

The approximate chi-squared value on one degree of freedom is

$$(U)^2/V = 787.92/51.62 = 15.26.$$

The statistic U has a negative sign because the exposure is protective — the observed number of vaccinated cases is less than would have been expected had vaccination been ineffective.

Exercise 18.2. Verify that, when there is only one case per stratum, the test becomes identical to the log rank test discussed in section 15.5.

This test was proposed by Mantel and Haenszel. They also proposed a way of calculating a nearly most likely value for θ. This is suggested by an algebraic rearrangement of the equation for the score:

$$\begin{aligned} U &= \sum (D_1^t - E_1^t) \\ &= \sum \frac{D_1^t H_0^t - D_0^t H_1^t}{N^t} \\ &= \sum Q^t - \sum R^t, \end{aligned}$$

where $Q^t = D_1^t H_0^t/N^t$ and $R^t = D_0^t H_1^t/N^t$. The usual estimate of the odds ratio in stratum t is Q^t/R^t, and this suggests estimating the common odds ratio, θ, by

$$\frac{Q^1 + Q^2 + \ldots}{R^1 + R^2 + \ldots} = \frac{Q}{R}.$$

When the true value of θ is close to 1, this *Mantel–Haenszel estimate* is almost as precise as the the most likely value of θ according to the hypergeometric likelihood. It can only be improved upon for odds ratios which differ substantially from one.

Exercise 18.3. Show that the Mantel–Haenszel estimate of the odds ratio for the data of Table 18.1 is 0.587.

Note that allowing for confounding by age has weakened the estimated protective effect of vaccination. This is now about 41% rather than 52% — a modest adjustment. This is in accord with the general experience that confounding only causes substantial modification of rate ratios in quite extreme circumstances.

The usefulness of the Mantel–Haenszel estimate in practice was limited by the fact that, rather surprisingly, no expression was available for its standard deviation until relatively recently. Several estimates have now been proposed, most of them rather awkward to calculate. For most practical purposes, a good estimate is provided by the same expression as for the cohort study version (Chapter 15):

$$S = \sqrt{\frac{V}{QR}}.$$

Exercise 18.4. For the data of Table 18.1, calculate the 90% confidence interval for the age-adjusted vaccine effect.

18.3 Improving efficiency by matching

In Exercise 16.2 we repeated the analysis of the leprosy study using a sample of 1000 controls drawn randomly from the healthy population, with only a modest loss in the precision of our estimate of the odds ratio. The position changes, however, when we stratify by age in the analysis.

Table 18.2 shows the way the simulated data lie. It is clear that the precision of the age-controlled odds ratio estimate will not be as good as we would have expected with more than 3 times as many controls as cases. The study has 238 controls for the 2 cases in the 0–4 year age group yet only 80 controls for the 53 cases in the 30–34 year age group.

With such a design, many controls are wasted and the efficiency of the study will be lower than it would be if the ratio of controls to cases were held constant within strata. This is called *matching*. If the study is carried out so as to achieve a constant ratio of cases to controls in broad groups it is called a *group* or *frequency* matching. If a set of matched controls are selected specifically for each case, it is called *individual matching*. Table 18.3 shows a simulated study in which the number of controls has been maintained at 4 times the number of cases in all age groups.

Exercise 18.5. For the data set out in Table 18.2, the values of Q, R, U, V are

Table 18.2. The simulated study stratified by age

	BCG scar			
	Cases		Controls	
Age	Absent	Present	Absent	Present
0–4	1	1	101	137
5–9	11	14	91	115
10–14	28	22	82	101
15–19	16	28	28	87
20–24	20	19	25	69
25–29	36	11	63	21
30–34	47	6	56	24

Table 18.3. A simulated group-matched study

	BCG scar			
	Cases		Controls	
Age	Absent	Present	Absent	Present
0–4	1	1	3	5
5–9	11	14	48	52
10–14	28	22	67	133
15–19	16	28	46	130
20–24	20	19	50	106
25–29	36	11	126	62
30–34	47	6	174	38

30.00, 51.57, 21.57, and 39.68. For Table 18.3 the corresponding values are 32.14, 56.54, 24.40, and 43.27. Compare the estimates, confidence intervals, and score tests for the two sets of data.

In practice, age is usually a very strong confounder and almost all case-control studies are matched for age. At one stage, simultaneously matching for as many other confounders as possible was frequently advocated. It is now clear that this is not a good idea, but matching is such an intuitively appealing idea to many epidemiologists that some discussion of the points for and against matching is of interest.

First it should be noted that an appreciable gain in precision is achieved only for a confounding variable which is very strongly related to the exposure of interest. For less strongly related confounders matching leads to only modest gains in precision while complicating the study design. More seriously, if a variable is matched in the design, the ability to examine the effect of that variable is lost since its distribution in the controls will match that in the cases rather than that in the study base. One must be confident

Table 18.4. Bias due to ignoring matching

	Cases		Controls		Odds
Stratum	Exposed	Unexposed	Exposed	Unexposed	ratio
1	89	11	80	20	2.0
2	67	33	50	50	2.0
3	33	67	20	80	2.0
Total	189	111	150	150	1.7

of the role and status of the variable before accepting such a limitation.

Secondly, much of the early popularity of matching stemmed from a misconception that variables matched in the design can be ignored in analysis, since differences between cases and controls could then not be attributable to these variables. It is now understood that this practice leads, in general, to incorrect estimates of odds ratios. This is demonstrated by Table 18.4. There are 100 cases and 100 controls in each stratum so that, overall, the cases and controls are matched with respect to stratum. However, despite the matching, the marginal odds ratio is 1.7 rather than 2.0, the value within strata. We have already warned of this behaviour of the odds ratio in section 15.7; even when confounding by age is removed by matching, the marginal odds ratio is not equal to the conditional (age-specific) odds ratios.

The bias that arises by ignoring matching in the analysis is always towards $\theta = 1$. The only circumstances under which it does not occur is when the matching variable is unrelated to exposures of interest. Only then may the matching be ignored, but in that case the variable is not a confounder and there would seem to be no purpose in matching for it in the first place. However, we shall see in the next section that there are reasons for matching other than for the efficient control of confounding. Some of these can lead to circumstances in which the matching can be ignored in analysis, but usually this is not the case.

Taken together, these two points lead us to a position where a matching variable *must* be regarded as a confounder and *must* be used in the analysis. From this it follows that estimates of the effects of all other exposures will be controlled for the matching variable. But this may not be what we want to do. For example, in perinatal epidemiology it may be appropriate in some analyses to consider birthweight as a confounder while for other analyses this may not be sensible. If the study is matched for birthweight at the design stage, analyses which seek to hold birthweight constant are easily carried out using stratified comparisons, but analyses which do not hold birthweight constant are much more difficult. Indeed they would be impossible without knowledge of the sampling fractions for drawing controls from the base within strata. These complexities are best avoided and

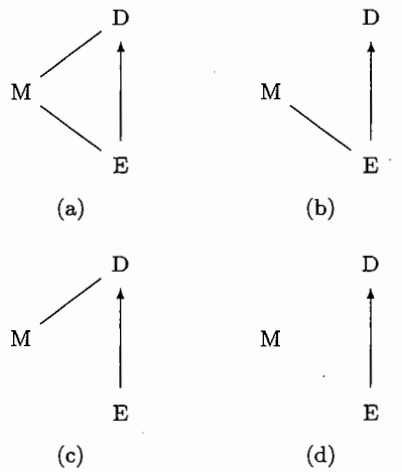

Fig. 18.1. To match or not to match?

matching for variables which may not be regarded as confounders for some questions is in general a mistake.

Finally, matching may actually *reduce* the efficiency of a study. This occurs when the matching variable is strongly related to the exposure, but not to disease risk (so that, again, it is not a confounder). This is called *overmatching*. It leads to a loss in efficiency because the effect of the matching is only to narrow the range of exposure studied. A good example would be a study of diet and some childhood illness using siblings of cases as controls. While such a study would be expected to yield the correct answer if properly analysed, it would be very inefficient — since siblings usually eat at the same table of the same prepared meals, the only information available for estimating the effects of interest will be from sibling pairs with discordant diets.

This discussion is summarized in Fig. 18.1. The letters D, E, and M refer to disease, exposures of interest, and matching variable respectively. Connecting lines indicate statistical relationship. Case (a) is the only one in which matching leads to a more precise estimate of the odds ratio. Case (b) is overmatching and leads to a *loss* of precision. In cases (a) and (b), the matching must be preserved in the analysis, whereas in cases (c) and (d) it may be ignored.

The above discussion tacitly assumes that controls are matched to cases in rather broad strata, such as 5- or 10-year age bands. It applies equally to individually matched studies; in principle there is no difference between

these options, although in practice the latter present rather more difficult analysis problems as a result of the very large number of nuisance parameters introduced by such fine stratification. These will be discussed in Chapter 19. Although matching must usually be preserved during analysis, it is not always necessary to preserve *individual* matching. If matching of controls to cases is only with respect to well defined, accurately measured variables then a coarser grouping at the analysis stage is both possible and acceptable. For example, if matching is only by age, analysis by 5- or 10-year bands will be quite satisfactory even if specific controls were drawn for each case. However, matching by characteristics such as neighbourhood or family does not allow later aggregation of strata.

18.4 Other reasons for matching

Matching is usually justified on the grounds of statistically efficient control for confounding. Close examination of this suggests that matching should be used as little as possible and only for variables, like age, which are strongly related to both disease and exposure and whose status is unequivocally that of confounder. However, a cursory review of the epidemiological literature shows that matching is used much more widely than this argument would support. This is because controls are often matched to cases for reasons which have nothing to do with control for confounding.

INCIDENCE DENSITY SAMPLING

One example is incidence density sampling, which is simply matching controls to cases with respect to *time* (date of occurrence). Although time may be a confounder (when both disease rate and exposure distribution in the study base vary during the study period), incidence density sampling is more usually employed for simple practical reasons. It will often be possible to ignore this matching in the analysis or, at most, to group coarsely on time.

DEFINING THE EXPOSURE WINDOW

Until this point we have assumed that each individual can be classified as exposed or unexposed and that this assignment holds for all time. However, many exposures in epidemiology vary over time, perhaps quite rapidly. When this is the case, it is necessary to specify the time period for assessing relevant exposure. This *exposure window* is usually clearly definable for cases, by working backwards from the point at which disease was first recognized, but comparable rules for controls can be difficult to specify. Things are much easier when one or more controls are matched to each single case with respect to time of diagnosis of the case; the time window used for assessing the relevant exposure of each case is carried over to the

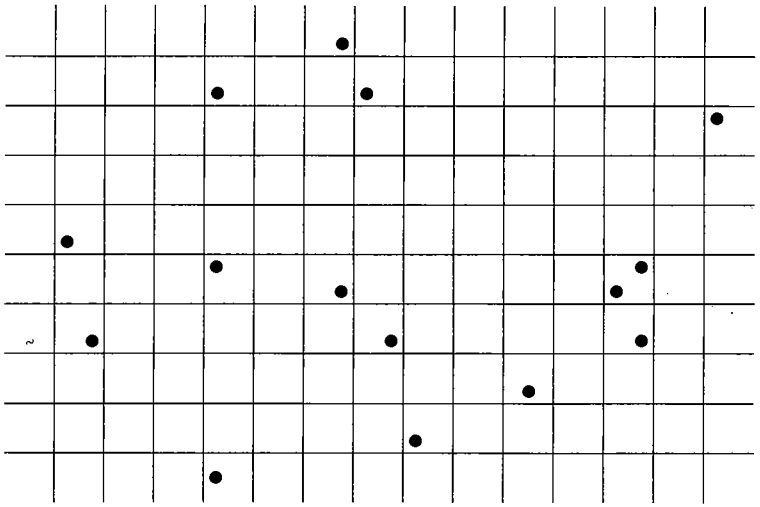

Fig. 18.2. Neighbourhood matching.

matched controls, thus ensuring comparability. We shall encounter a good example of this in Chapter 19.

AVOIDING SELECTION BIAS

Another example is where controls are matched to cases in order to minimize selection bias. This is usually done either because the study base has not been precisely defined or because there is no accurate way of sampling it. For example, in a geographically based study selection bias may be caused by the lack of an accurate population register of the study area. Unfortunately, construction and maintenance of such registers is enormously costly and will rarely be feasible for a single case-control study. However, if the study is closely matched, better sampling may be possible. Fig. 18.2 illustrates this for a geographically based study, divided by the grid into small *neighbourhoods*. The dots represent cases occurring during the study period. A study which matched for neighbourhood would sample controls only from those neighbourhoods in which a case occurred and it would only be necessary to construct lists of eligible controls for these. If neighbourhoods are sufficiently small this involves little work. Of course, the definition of neighbourhood does not have to be in terms of a regular grid for this argument to apply. A similar argument justifies drawing controls from the list of patients of the family doctor of each case.

Solutions to the exercises

18.1 The number of exposed cases, D_1^1, is 1 and the expected number under the null hypothesis is

$$E_1^1 = 2 \times \frac{11720}{11720 + 7594} = 1.21$$

so that the contribution to U is $(1 - 1.21) = -0.21$. The contribution to V is

$$\frac{2 \times 19312 \times 7594 \times 11720}{(19314)^2 \times 19313} = 0.48.$$

18.2 The expression for the 'expected' number of exposed cases in each stratum, E_1^t, is identical to that given in section 15.5. Thus, the score statistics, U, are identical. When there is only one case per stratum, $D^t = 1$ and $H^t = N^t - 1$ so that the contribution of stratum t to V is

$$V^t = \frac{(N^t - 1)N_0^t N_1^t}{(N^t)^2 (N^t - 1)} = \frac{N_0^t N_1^t}{(N^t)^2},$$

which is identical to our previous expression. When using the log rank test with tied event occurrence times (so that $D^t > 1$), the variance formula given in this chapter should be used.

18.3 The first contribution to the numerator (top) and denominator (bottom) of the Mantel–Haenszel estimate are as follows:

$$Q^1 = \frac{1 \times 7593}{1 + 1 + 7593 + 11719}, \qquad R^1 = \frac{1 \times 11719}{1 + 1 + 7593 + 11719}.$$

Continuing the calculation, we get:

Age	Q^t	R^t
0–4	0.39	0.61
5–9	5.76	6.46
10–14	9.34	16.01
15–19	5.96	12.53
20–24	5.74	13.86
25–29	7.95	9.70
30–34	4.82	8.88
Total	39.96	68.05

Note that the ratio Q/R for each row gives the odds ratios calculated in the previous exercise. The Mantel–Haenszel estimate is $39.96/68.05 = 0.587$.

18.4 V is given in the text following the first exercise as 51.62 and Q and R were calculated in the second exercise to be 39.96 and 68.05 respectively. Using the formula $S = \sqrt{V/(QR)}$,

$$S = \sqrt{\frac{51.62}{39.96 \times 68.05}} = 0.138$$

The error factor for 90% confidence limits is $\exp(1.645 \times 0.138) = 1.255$ so that the confidence limits for the odds ratio controlled for age are $0.587/1.255 = 0.47$ (lower limit) and $0.587 \times 1.255 = 0.74$ (upper limit).

18.5 The analysis of the two sets of data yields the following results:

	Table 18.2	Table 18.3
Estimate (θ)	0.582	0.568
S ($\log(\theta)$)	0.160	0.154
Error factor	1.301	1.289
Lower 90% limit	0.447	0.441
Upper 90% limit	0.757	0.732
$(U)^2/V$	11.73	13.76

In this case the increase in precision achieved by matching is not great.