x-rays in infancy—fourth survey in 20 years. *J. Natl. Cancer Inst.* 1975;55:519–530.
Kleinbaum, D. G., and Kupper, L. L. *Applied Regression Analysis and Other Multivariate Methods.* North Scituate, MA: Duxbury Press, 1978. P. 243.
Lubin, J. H. An empirical evaluation of the use of conditional and unconditional likelihoods for case-control data. *Biometrika* 1981;68:567–571.
Mann, J. I., Inman, W. H. W., and Thorogood, M. Oral contraceptive use in older women and fatal myocardial infarction. *Br. Med. J.* 1968;2:193–199.
Mantel, N., and Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* 1959;22:719–748.
Mantel, N., Brown, C., and Byar, D. P. Tests for homogeneity of effect in an epidemiologic investigation. *Am. J. Epidemiol.* 1977;106:125–129.
McKinlay, S. M. The effect of nonzero second-order interaction on combined estimators of the odds ratio. *Biometrika* 1978;65:191–202.
Miettinen, O. S. Standardization of risk ratios. *Am. J. Epidemiol.* 1972;96:383–388.
Miettinen, O. S., and Neff, R. K. Computer processing of epidemiologic data. *Hart Bull.* 1971;2:98–103.
Nurminen, N. Asymptotic efficiency of general noniterative estimators of common relative risk. *Biometrika* 1981;68:525–530.
Robins, J. M., Breslow, N., and Greenland, S. Estimators of the Mantel-Haenszel variance consistent in both sparse data and large strata limiting models. *Biometrics* 1986;42:311–323.
Rothman, K. J. Spermicide use and Down's syndrome. *Am. J. Public Health* 1982;72:399–401.
Rothman, K. J., and Monson, R. R. Survival in trigeminal neuralgia. *J. Chron. Dis.* 1973;26:303–309.
Rothman, K. J., and Boice, J. D. *Epidemiologic Analysis with a Programmable Calculator.* Brookline, MA: Epidemiology Resources, 1982. (First edition published by U.S. Government Printing Office, Washington D.C., NIH Publication No. 79-1649, June, 1979.)
Tarone, R. E. On summary estimators of relative risk. *J. Chron. Dis.* 1981;34:463–468.
Thomas, D. G. Exact and asymptotic methods for the combination of 2 × 2 tables. *Computers and Biomedical Research* 1975;8:423–446.
University Group Diabetes Program. A study of the effects of hypoglycemic agents on vascular complications in patients with adult onset diabetes. *Diabetes* 1970;19(Suppl. 2):747–830.
Walker, A. M. Small sample properties of some estimators of a common hazard ratio. *Appl. Stat.* 1985;34:42–48.
Woolf, B. On estimating the relation between blood group and disease. *Ann. Hum. Genet.* 1954;19:251–253.
Zelen, M. The analysis of several 2 × 2 contingency tables. *Biometrika* 1971;58:129–137.

13. MATCHING

Matching refers to the selection of a comparison series—unexposed subjects in a follow-up study or controls in a case-control study—that is identical, or nearly so, to the index series with respect to one or more potentially confounding factors. The mechanics of the matching may be performed subject by subject, which is described as *individual matching*, or for groups of subjects, which is described as *frequency matching*. The general principles that apply to matched data are identical for individually matched or frequency matched data.

PRINCIPLES OF MATCHING

The topic of matching in epidemiology is beguiling: What at first seems clear is seductively deceptive. Whereas the clarity of an analysis in which confounding has been securely prevented by perfect matching of the compared series seems indubitable and impossible to misinterpret, the intuitive foundation for this cogency attained by matching is a surprisingly shaky structure that does not always support the conclusions that are apt to be drawn. The difficulty is that our intuition about matching springs from knowledge of experiments or follow-up studies, whereas matching is most often applied in case-control studies, which differ enough from follow-up studies to make the implications of matching different and counterintuitive.

Whereas the traditional view, stemming from an understanding based on follow-up studies, has been that matching enhances validity, in case-control studies the effectiveness of matching as a methodologic tool derives from its effect on study efficiency, not on validity. Indeed, for case-control studies it would be more accurate to state that matching introduces confounding rather than that it prevents confounding.

The different implications of matching for follow-up and case-control studies are easy to demonstrate. Consider a source population of 2,000,000 individuals, distributed by exposure and sex as indicated in Table 13-1. Both the exposure and male gender are risk factors for the disease: For the exposure the relative risk is 10, and for males relative to females it is 5. There is also substantial confounding, since 90 percent of the exposed individuals are male and only 10 percent of the unexposed are male. The crude relative risk in the source population, comparing exposed with unexposed, is 32.9, considerably different from the unconfounded value of 10.

Now consider what happens if a follow-up study is planned by drawing the exposed cohort from the exposed source population and matching the unexposed cohort to the exposed cohort for sex. Suppose 10 percent of the exposed source population were included in the follow-up study; if these subjects were selected independently of gender, we would have approximately 90,000 males and 10,000 females in the exposed cohort. A

Table 13-1. Hypothetical source population of 2,000,000 people, in which exposure increases risk 10-fold, and males have five times the risk of females, and exposure is strongly associated with male gender

	Males (1,000,000)		Females (1,000,000)	
	Exposed (900,000)	Unexposed (100,000)	Exposed (100,000)	Unexposed (900,000)
One-year risk	0.005	0.0005	0.001	0.0001
No. cases in one year	4500	50	100	90

$$\text{Crude relative risk} = \frac{(4500 + 100)/1{,}000{,}000}{(50 + 90)/1{,}000{,}000} = 32.9$$

comparison group of unexposed subjects would be drawn from the 1,000,000 unexposed individuals in the source population. If the comparison group were drawn, like the exposed group, independently of gender, the follow-up study would have the same confounding as exists in the source population (apart from sampling variability), since the follow-up study would then be a simple 10 percent sample of the source population. It would be possible, however, to assemble the unexposed cohort so that the proportion of males in it was identical to that in the exposed cohort. The purpose of matching the unexposed cohort to the exposed group by sex is to prevent confounding by sex. Of the 100,000 unexposed males in the source population, 90,000 would be in a matched comparison group, corresponding to the 90,000 exposed males in the study. Of the 900,000 unexposed females, 10,000 would be selected to match the 10,000 exposed females.

The "expected" results from the matched cohort study described here are indicated in Table 13-2. The expected relative risk in the study population is 10 for males and 10 for females and is also 10 in the crude data for the study. The matching has apparently accomplished its purpose: There is no confounding by sex, since sex is unrelated to exposure in the study population because of matching.

The situation differs considerably, however, if a case-control study is conducted instead. Consider a case-control study based on the total of 4740 cases that occur in the source population during one year. Of these cases, 4550 are male. Suppose, then, that 4740 controls were selected from the source population, matched to the cases by gender, so that 4550 of the controls are male. Of the 4550 male controls, we expect about 90 percent, or 4095, to be exposed, since 90 percent of the males in the source population are exposed. Of the 190 female controls, we expect about 10 percent, or 19, to be exposed, corresponding to the 10 percent of females exposed in the source population. For the control series as a whole, the expected number of exposed subjects is 4095 + 19 = 4114 of a total of

Table 13-2. Expectation of the results of a matched one-year follow-up study of 100,000 exposed and 100,000 unexposed subjects drawn from the source population described in table 13-1

	Males		Females		Total	
	Exposed	Unexposed	Exposed	Unexposed	Exposed	Unexposed
Cases	450	45	10	1	460	46
Total	90,000	90,000	10,000	10,000	100,000	100,000
	$\hat{RR} = 10$		$\hat{RR} = 10$		Crude $\hat{RR} = 10$	

4740. For the cases, 4500 + 100 = 4600 of the 4740 would be exposed. The crude estimate of effect, based on the odds ratio from the crude data, is

$$\text{Crude relative risk} \doteq \frac{(4600)\,(626)}{(4114)\,(140)} = 5.0$$

which is a substantial underestimate of the unconfounded effect of the exposure. Interestingly, the case-control data give the correct result, $\hat{RR} = 10$, if the data are stratified into male and female strata (Table 13-3). The discrepancy between the crude results and the stratum-specific results in Table 13-3 is a manifestation of confounding by sex (note that the sex-specific effect estimates are identical to one another and distinctly different from the crude estimate). This confounding is not a reflection of the original confounding by sex in the source population but rather a confounding that was introduced into the study by the matching process. In case-control studies, matching on factors associated with exposure builds confounding into the data, whether or not there was confounding in the source population. If there is confounding initially in the source population, as there was in the example, the process of matching will substitute a new confounding structure in place of the initial one. The confounding introduced by matching is generally in the direction of a bias toward the null value of effect, whatever the nature of the confounding in the source population. In the example, the strong positive confounding (positive indicating a bias in the same direction as the effect) in the source population was replaced by strong negative confounding (negative indicating a bias in the direction opposite to that of the effect) in the case-control data.

Why does matching in a case-control study introduce confounding? The purpose of the control series in a case-control study is to provide an estimate of the person-time distribution for the exposed population relative to the unexposed population in the source population of cases. If controls are selected to match the cases for a factor that is correlated with the exposure, then the crude exposure proportion in controls is distorted in

Table 13-3. Expectation of the results of a case-control study of 4740 cases and 4740 matched controls drawn from the source population described in table 13-1

	Males			Females			Total		
	Exposed	Unexposed	Total	Exposed	Unexposed	Total	Exposed	Unexposed	Total
Cases	4500	50	4550	100	90	190	4600	140	4740
Controls	4095	455	4550	19	171	190	4114	626	4740
	$\hat{RR} = 10$			$\hat{RR} = 10$			$\hat{RR} = 5$		

the direction of similarity to that of the cases. If the matching factor were perfectly correlated with the exposure, the exposure distribution of controls would be identical to that of cases, and the crude relative risk estimate would be 1.0, since controls are chosen to be identical to cases with respect to the matching factor. Interestingly, the bias of the effect estimate toward the null value does not depend on the direction of the correlation between the exposure and the matching factor; as long as there is a nonzero correlation, positive or negative, the crude exposure distribution among controls will be distorted in the direction of similarity to that of cases. A perfect negative correlation between the matching factor and the exposure will still lead to identical exposure distributions for cases and controls and a crude relative risk estimate of 1.0 because each control is matched to the identical value of the matching factor of the case, guaranteeing identity for the exposure variable as well.

If the matching factor happens to be uncorrelated with the exposure, then matching does not influence the exposure distribution of the controls, and therefore no bias is introduced by matching. Because matching is ostensibly motivated by the need to control confounding by the matching factor(s), one would generally expect some correlation to exist between the matching factor(s) and the exposure. If the correlation is zero, the matching factor was not confounding in the first place, since a confounding factor must be associated with both the exposure and the disease.

It seems that matching, although intended to control confounding, does not attain that objective in case-control studies. Apparently, it merely accomplishes the substitution of a new confounding structure for the old one. In fact, matching can even introduce confounding where none previously existed: If the matching factor is unrelated to disease in the source population, ordinarily it would not be a confounder; however, if it is correlated with the exposure, it will become a confounder after matching for it in a case-control study. This situation is illustrated in Table 13-4, in which the exposure has an effect corresponding to a relative risk of 5.6, and there is no confounding in the source population; however, if the cases are used as the basis for a case-control study, and a control series is matched to the cases by gender, the expected value for the crude estimate of effect from the case-control study is 2.1 rather than the correct value of 5.6. In the source population sex is uncorrelated with disease among the unexposed, the prevalence of disease being 2 in 1000 for both unexposed males and unexposed females. Sex is strongly correlated with exposure, however. In the case-control study, sex is confounding because it was a matching factor that was correlated with exposure. Despite the absence of correlation between sex and disease among unexposed in the source population, a correlation between sex and disease among unexposed is introduced into the case-control data by matching. The result is a crude estimate of effect, 2.1, that seriously underestimates the correct value of 5.6.

Table 13-4. Prevalence state of a source population with no confounding by sex and a case-control study drawn from the source population illustrating confounding introduced by matching for sex.

Source population

	Males			Females			Totals	
	Exposed	Unexposed		Exposed	Unexposed		Exposed	Unexposed
Disease	999	20		111	180		1110	200
No disease	89,001	9980		9889	89,820		98,890	99,800
Totals	90,000	10,000		10,000	90,000		100,000	100,000
		$\hat{RR} = 5.6$			$\hat{RR} = 5.6$		Crude $\hat{RR} = 5.6$	

Case-control study drawn from the source population and matched on sex

	Males			Females			Totals		
	Exposed	Unexposed	Total	Exposed	Unexposed	Total	Exposed	Unexposed	Total
Cases	999	20	1019	111	180	291	1110	200	1310
Controls	916	103	1019	29	262	291	945	365	1310
Totals	1915	123	2038	140	342	482	2055	565	2620
		$\hat{RR} = 5.6$			$\hat{RR} = 5.6$		Crude $\hat{RR} = 2.1$		

The confounding introduced by matching in a case-control study is by no means irremediable. Notice that in Tables 13-3 and 13-4 the stratum-specific estimates of effect are valid; the confounding can be removed by a stratified analysis to arrive at a pooled estimate of effect after stratifying by the matching factor(s). Table 13-4 illustrates the need for an analysis to remove confounding by the matching factors, since matching may cause confounding even when none was originally present. In Table 13-4, the selection criterion used in matching controls makes the control series unrepresentative of the source population with regard to exposure; this would lead to a selection bias but for the fact that it can be controlled in the analysis and can be therefore viewed as confounding.

In a follow-up study that compares risks, no additional action is required in the analysis to control confounding by the matching factors; the process of matching has already eliminated any confounding by the matching factors. In contrast, matching in a case-control study requires further control of confounding by the matching factors in the analysis even if the matching factors were not confounding in the source population, provided that the matching factors are correlated with the exposure. What accounts for this discrepancy? In a follow-up study, matching is undertaken without regard to disease status, which is unknown at the start of follow-up, therefore preventing bias. In a case-control study, on the other hand, matching involves the specification of both the exposure and the disease status and leads to conditional associations between the matching factor(s) and both exposure and disease, thereby resulting in bias. In a case-control study, if the matching factors are not correlated with the exposure, no confounding is introduced by matching; in this situation there could not have been confounding in the source population to begin with, so the matching was unnecessary.

It is reasonable to ask why one would consider matching at all in case-control studies, since it does not accomplish its intended objective of preventing confounding. The utility of matching in case-control studies derives not from its ability to prevent confounding but from the enhanced efficiency that it affords for the control of confounding. In Table 13-3, the male and female strata each have an equal number of cases and controls because of the matched design. If 4740 controls were selected without matching, half would be male and half would be female. There would thus be a great excess of female controls, since 2370 is an unnecessarily large number of controls for 190 cases; the total amount of information does not increase substantially after five or six controls per case (see Fig. 8-1), and therefore the information collected on so many females is partially wasted. On the other hand, there would be only 2370 male controls for the 4550 male cases. It is generally inefficient to have strata in which the ratio of controls to cases varies substantially on either side of unity. The extreme form of such inefficiency occurs when there are many individual strata with one or more cases and no control subjects (control/case ratio

= 0) and other strata with one or more controls and no cases (control/case ratio = infinity). Such strata provide no information in a stratified analysis. If matching is used in the selection of controls, however, there will be fewer uninformative strata in a stratified analysis than there would have been in such an analysis without matching: A fixed number of matched controls for each case will provide an extremely efficient stratified analysis. The improved efficiency will be manifest in narrower confidence limits about the point estimate than would otherwise be obtainable. Matching in case-control studies can thus be considered a means of providing a more efficient stratified analysis rather than a direct means of preventing confounding. Stratification (or an equivalent multivariate approach) will be necessary to control confounding with or without matching, but matching makes the stratification more efficient.

The efficiency that matching provides in the analysis of case-control data comes at a substantial cost. One part of the cost is a research limitation: If a factor has been matched in a case-control study, it is no longer possible to estimate the effect of that factor, since its distribution is forced to be identical for cases and controls. Consequently, matching factors cannot be the objects of inquiry in a case-control study (except as effect modifiers—see pages 279–282, Evaluation of Effect Modification with Matched Data). Another cost is the added analytic complexity required to control confounding by factors that have not been matched. It is possible to control simultaneously for both matched and unmatched factors but usually only through specialized analyses, usually multivariate models. Conducting these analyses poses no serious difficulties in view of the growing availability of computers, but the investigator is forced to depend on computers and computer programs to analyze data that might otherwise have been analyzed in a more straightforward way.

A further cost involved with individual matching is the literal expense entailed in the process of choosing control subjects with the same distribution of matching factors found in the case series. If several factors are being matched, many potential control subjects must typically be scanned to find one that has the same characteristics as the case. Whereas this arduous process may lead to a statistically efficient analysis, it improves efficiency only at considerable expense.

If the efficiency of a study is judged from the point of view of the amount of information per subject studied (size efficiency), matching can be viewed as a means of improving study efficiency. Alternatively, if efficiency is judged as the amount of information per unit of cost involved in obtaining that information (cost efficiency), matching may paradoxically have the opposite effect of decreasing study efficiency, since the effort expended in finding matched subjects could be spent simply in gathering information for a greater number of unmatched subjects. With or without matching, confounding would have to be controlled in the data analysis. With matching, a stratified analysis would be more size efficient, but without it the

resources for data collection can increase the number of subjects, thereby improving cost efficiency. Since cost efficiency is a more fundamental concern to an investigator than size efficiency, the apparent efficiency gains from matching may be illusory.

Thus the beneficial effect of matching on study efficiency, which is the primary reason for employing matching, appears to be ephemeral. Indeed, the decision to match subjects can result in less overall information, as measured by the width of the confidence interval for the effect measure, than would have been obtained without matching if the expense of matching reduces the total number of study subjects. A wider appreciation for the costs that matching imposes and the often meager advantages it offers would presumably persuade epidemiologists to avoid the technique in many settings in which matching is routinely used. Since the intended goal is to control confounding, and this goal is attainable only by proper analysis regardless of whether matching is employed, the routine use of matching is seldom justified.

Nevertheless, there are some situations in which matching is desirable or even necessary. If the process of obtaining the information from the study subjects is expensive, it is desirable to optimize the amount of information obtained per subject. For example, if exposure information in a case-control study involves an expensive laboratory test run on blood samples, the investigator would want the information from each subject to contribute as much as possible. As long as the expense of ascertaining matched controls is small compared with the expense of obtaining the exposure information from each subject, it is preferable to plan for a stratified analysis in which the stratification does not lead to loss of information, that is, it is desirable to match controls during subject selection so that there will be a uniform ratio of controls to cases in the stratified analysis. If no confounding is anticipated, of course, there is no need to match; for example, restriction of both series might prevent confounding without the need for stratification or matching. If confounding is likely, however, matching will ensure that control of confounding in the analysis will not lose information that has been expensively obtained. The essential difference that makes matching attractive in this situation is the high price of expanding the study size; when additional subjects are expensive to obtain, it is worthwhile to pay the cost of matching to take full advantage of the information that is collected. In such a situation, matching serves both size efficiency and cost efficiency.

Sometimes the control of confounding in the analysis is not possible unless matching has prepared the way to do so. Imagine a potential confounding factor that is measured on a nominal scale with many categories; examples would be variables such as neighborhood, sibship, and occupation. Controlling sibship would be impossible unless sibling controls had been selected for the cases, that is, matching on sibship is required to control for it. These variables are distinguished from other nominal scale

variables such as sex by their multitude of categories, ensuring that one or very few subjects will fall into each category. Without matching, most strata in a stratified analysis would have only one subject, either a case or a control, and no information about effect unless control subjects had been matched to the cases for the value of the factor in question. Continuous variables such as age also have a multitude of values, but the values are easily combined by grouping, avoiding the fundamental problem. If the categories of a nominal scale variable could be combined in a reasonable way, the need for matching could be avoided. Methods to achieve this have been proposed [for example, see Miettinen, 1976], but they require a multivariate analysis as a preliminary step to the stratified analysis. Matching for nominal scale variables with many categories ensures that, after stratification by the potentially confounding factor, each case will have one or more matched controls for comparison.

A fundamental problem with stratified analysis is the inability to control confounding by several factors simultaneously. Control of each additional factor involves spreading the existing strata over a new dimension; the total number of strata required becomes exponentially large as the number of stratification variables increases. For studies with many confounding factors, the number of strata in a stratified analysis that controls all factors simultaneously may be so large that the situation mimics that in which there is a nominal scale confounder with a multitude of categories: There may be one or very few subjects per stratum and hardly any comparative information about the effect in any strata. If a large number of confounding factors is anticipated, matching may be desirable to ensure an informative stratified analysis. On the other hand, it is not absolutely necessary to match unless there are nominal scale variables with many categories, since a multivariate analysis can cope with confounding by many factors simultaneously even in situations in which stratification fails. Even multivariate analysis, however, is inadequate to control confounding by nominal scale variables with a large number of possible values unless matching has provided the necessary comparative information within categories.

We can summarize the utility of matching in case-control studies as follows: Matching is a useful means for improving study efficiency, in terms of the amount of information per subject studied, if the amount of information obtainable from the more efficient analysis exceeds the amount of information obtainable simply by studying more subjects without matching. Matching is indicated for potentially confounding factors that are measured on a nominal scale with many categories or when the number of potentially confounding variables is so great that stratification would spread the subjects too thinly over the strata. Multivariate analysis is a reasonable alternative in the latter situation; it would be feasible even without matching. Even multivariate analysis, however, is infeasible to control confounding by a nominal scale factor with many categories, unless matching is employed.

Overmatching

A term often used in reference to matched studies is *overmatching*. The interpretation of this term has changed with a sharper understanding of the principles that underlie matched studies. Originally, the term overmatching was used to refer to a loss of validity in a case-control study stemming from a control group that was so closely matched to the case group that the exposure distributions differed very little. This original interpretation for overmatching was based on a faulty analysis that failed to correct for confounding. On proper analysis, no validity problem whatsoever is introduced by matching. Note that in a follow-up study with matching even the crude analysis is valid, so that overmatching was never seen as a problem for follow-up studies. We have seen that indeed a validity problem does exist from matching in a case-control study if the crude data are used for inference. This problem disappears, however, if stratification by the matching factors is employed in the analysis.

The modern interpretation of overmatching relates to study efficiency rather than validity. Consider an individually matched case-control study with one control matched to each case. Each stratum in the analysis will consist of one case and one control unless some strata can be combined. A stratum cannot contribute information to a case-control analysis if any marginal total in the 2×2 table is equal to zero. If a case and a single matched control are either both exposed or both unexposed, one margin of the 2×2 table will be zero and that pair of subjects will not contribute any information to the analysis. If several controls are matched to a single case and all the controls have the same exposure value as the case, all exposed or all unexposed, the resulting zero margin likewise signals that the matched set of controls and case will not contribute to the analysis. Since matching is intended to select controls identical to the index case with respect to correlates of exposure, typically the information from many subjects is "lost" in a matched analysis. Obviously the loss of information detracts from study efficiency, reducing both information per subject studied and information per dollar spent. Matching has the net effect of increasing study efficiency only because stratified analysis in the absence of matching is ordinarily even less efficient than stratified analysis with matching. Recall, however, that matching in a case-control study can introduce confounding even if none exists in the source population, if the matching factor is correlated with the exposure but not with the disease. In such an instance, matching decreases study efficiency by locking the investigator into an analysis stratified by the matching factor, which will inevitably lose information on the matched sets with completely concordant exposure histories, whereas without matching a much more efficient crude analysis could have been used. Since the matching was not necessary in the first place and has the effect of impairing study efficiency relative to the type of analysis that could have been performed without matching, matching in this situation can properly be described as overmatching.

Overmatching is thus understood to be matching that causes a loss of information in the analysis because the resulting stratified analysis would have been unnecessary without matching. The extent to which information is lost by matching depends on the degree of correlation between the matching factor and the exposure. A strong correlate of exposure that has no relation to disease is the worst factor to match for, since it will lead to relatively few informative strata in the analysis with no offsetting gain. Consider, for example, a study of the relation between coffee drinking and cancer of the bladder; suppose matching for consumption of powdered cream-substitutes were considered along with matching for a set of other factors. Since this factor is a strong correlate of coffee consumption, many of the individual strata in the matched analysis will be completely concordant for coffee drinking and will not contribute to the analysis; that is, for many of the cases, controls matched to that case will be classified identically to the case with regard to coffee drinking simply because of matching for consumption of powdered cream-substitutes. If powdered cream-substitutes have no relation to bladder cancer, nothing is accomplished by the matching. Though no validity problem exists, the matching is counterproductive and can consequently be considered overmatching.

Matching on a risk factor that is not correlated with the exposure under study will not lead to an increased correlation of exposure histories for cases and controls. Such matching could nevertheless be considered overmatching because it adversely affects cost efficiency although it does not affect size efficiency. (Similarly, matching for any factor that is merely a consequence of disease can also be considered overmatching.) On the other hand, overmatching from a factor that is associated with exposure but not with the disease, such as indicators of opportunity for exposure [Poole, 1986], will reduce both cost efficiency and size efficiency, that is, an investigator will spend more to obtain information from the same number of subjects as he could have obtained without matching on the factor and will obtain less information per subject after having spent more. These losses in efficiency are suffered to control a factor that was not confounding anyway.

If a factor is a weak risk factor and a strong correlate of the exposure, it will be a weak confounder; matching for such a factor will involve a relatively large loss of information compared with a crude analysis because of the strong correlation with exposure. A crude analysis is no longer a proper alternative, however, if the factor is a genuine confounding factor. A reasonable alternative to matching of a confounding factor is a stratified analysis without matching. Matching theoretically improves efficiency by stabilizing the control-case ratio in the analysis, but it reduces efficiency by causing the loss of information in some strata in which the exposure information is concordant. If the elementary strata corresponding to each matched set have a reasonably large number of controls, complete concordance is unlikely; on the other hand, such concordance is very likely

for matched pairs. If elementary strata can be combined in the analysis, a possibility when there are only a few matching factors with a modest number of categories, it is much less likely that there will be zero margins for the 2 × 2 tables in the analysis. The likelihood is further reduced if the matching factors are not strongly correlated with the exposure, although it should be remembered that the confounding that prompts matching depends on the magnitude of the association between the potential confounding factor and the exposure: With no association, there is no confounding. If it is thought that a study design would lead to many elementary strata with zero margins, then the value of matching in stabilizing the case-control ratio to improve study efficiency must be weighed against the loss of information from concordant exposure histories. It may be considered a form of overmatching to match on a weak confounding factor that is a strong correlate of exposure, since the matching itself is expensive and can lead to a less efficient analysis than the alternative of stratification without matching. The primary way to improve study efficiency when considering matching for a strong correlate of exposure is to increase the ratio of controls to cases, thereby decreasing the likelihood of a zero margin in the 2 × 2 table corresponding to each matched set.

Matching on Indicators of Information Quality

Another reason that matching is sometimes employed is to achieve comparability in the quality of information collected. A typical situation in which such matching might be undertaken is a case-control study in which some or all of the cases have already died, and surrogates must be interviewed for exposure and confounder information. In principle, controls for dead cases should be living, since they constitute a sample from the source population that gave rise to the cases. In practice, since surrogate interview data is usually presumed to differ in quality from interview data obtained directly from the subject, many investigators prefer to match dead controls to dead cases. It is not clear, however, that matching on information quality is justifiable. Whereas using dead controls can be justified in "proportional mortality" studies essentially as a convenience (see Chapter 6), there is no certainty that matching on information quality reduces overall bias. Many of the assumptions about the quality of surrogate data, for example, are unproved [Gordis, 1982]. Furthermore, comparability of information quality still allows bias from nondifferential misclassification, which is more severe in matched than in unmatched studies [Greenland, 1982], and can be more severe than the bias due to differential misclassification arising from noncomparability [Greenland and Robins, 1985b].

To summarize, the intricacies of matching in case-control studies and the relation of matching to confounding and study efficiency are much more complicated than one might at first suppose. Matching has often been employed when simpler and cheaper alternatives would have been

preferable. Matching is clearly indicated only in sharply defined circumstances. In many study situations, the decision rests on cost and efficiency considerations that border on the imponderable.

HED CASE-CONTROL ANALYSIS

The most important point in the analysis of matched case-control data is that matching introduces a bias in the crude estimate of effect toward the null value if the matching factor is correlated either positively or negatively with exposure, conditional on disease status. This bias may be viewed as a type of confounding, since it is present in the crude data, but it can be completely removed by stratifying by the matching factors. Therefore, the main task in a matched case-control analysis is to stratify by the matching factors.

Since stratification has already been discussed, there would be no need to elaborate further on matched case-control analysis but for one special feature of these analyses: Often the matching factor or factors have so many possible categories that the stratified analysis consists of one stratum for every case in the study. This feature introduces no new analytic concepts into the stratified analysis beyond those discussed in Chapter 12, but it does often lead to analyses with dozens or hundreds of strata. The formulas of Chapter 12 become tedious to apply by hand if the number of strata is large, but the formulas can be simplified for matched data so that their application with pencil and paper is not arduous even with thousands of strata.

If stratification could be accomplished without creating a large number of strata, a study with matching could be analyzed using an ordinary stratified approach. For example, if subjects are matched only for age and sex, there is no need to conduct a specialized "matched analysis" that amounts to creating individual strata for each matched set of subjects. It is sufficient to consider age and sex as confounding factors that need to be controlled in the analysis and to create only the few strata for age and sex that would have been necessary had no matching been undertaken in subject selection. Additional confounding factors can be easily controlled in such an analysis, even if they are not matching factors, by further stratification or multivariate analysis. Frequency matching is always handled using a "nonmatched" analysis, that is, using the usual analytic techniques to control confounding. There is no special principle underlying the methods of a matched analysis. The need for a matched analysis is purely a practical one, stemming from the need to define strata in such a way that a large number of strata is inevitable, as is the case with an analysis in which a nominal scale variable with many categories is one of the matching variables. When such a variable is confounding, individual matching is needed to permit the control of confounding. In other situations, however, frequency matching or no matching at all is a better design option, since it is usually more cost efficient. Even if individual matching is employed, unless the number of categories in the analysis is inevitably large relative to the number of cases, there is no compulsion to use the methods of individually matched analysis as long as the matching factors are all controlled in the analysis.

Point Estimation of the Relative Risk (Odds Ratio) from Matched Case-Control Data

As usual for case-control data, the odds ratio, being an estimate of the incidence rate ratio or relative risk, is the measure of interest. Either the maximum likelihood or the Mantel-Haenszel approach may be used for estimation. The Mantel-Haenszel approach is simpler, but the maximum likelihood approach is not as complicated for matched data as it is for the usual stratified analysis.

Maximum likelihood estimation of the odds ratio in a stratified analysis can be "conditional" on both margins of the 2 × 2 tables or "unconditional," which means conditional on only one margin of the 2 × 2 table. The two approaches give nearly identical results except when the average number of subjects per stratum is small, in which case the unconditional approach can be substantially biased and should not be used [Breslow, 1981; Lubin, 1981]. Matched analyses are the extreme form of stratified analysis in the sense of having the fewest possible subjects per stratum. One case and one control per stratum is the minimum requirement, but studies in which all the strata are matched pairs can nevertheless be extremely informative. For matched analyses, the applicable likelihood methods are those based on the conditional likelihood.

POINT ESTIMATION OF RELATIVE RISK FROM MATCHED CASE-CONTROL PAIRS

When a single control is matched individually to each case, the elementary strata in the analysis are 2 × 2 tables with only two subjects. For a dichotomous exposure, only four possible exposure patterns exist for the two subjects: both exposed, both unexposed, case exposed and control unexposed, and case unexposed and control exposed. These four exposure patterns are shown in Table 13-5. Note that when the exposure history is identical for the case and the control, there is a marginal total equal to zero in the 2 × 2 table. The first and last of the 2 × 2 tables in Table 13-5, A and D, have a zero marginal total and consequently do not contribute to either estimation or statistical hypothesis testing.

The conditional maximum likelihood estimate of the odds ratio is simply the frequency of matched sets of type B divided by the frequency of sets of type C, that is, the ratio of the number of discordant pairs in which the case is exposed to the number of discordant pairs in which the control is exposed. This estimator can be derived easily as follows: If the odds

Table 13-5. Possible patterns of exposure for a case and a single matched control

	A			B			C			D		
	E	U	T	E	U	T	E	U	T	E	U	T
Case	1	0	1	1	0	1	0	1	1	0	1	1
Control	1	0	1	0	1	1	1	0	1	0	1	1
Totals	2	0	2	1	1	2	1	1	2	0	2	2

E = exposed; U = unexposed; T = total.

ratio is designated as OR, then from the noncentral hypergeometric distribution (see Chap. 11) the probability of a 2 × 2 table of type B is OR/(OR + 1) and the probability of a table of type C is 1/(OR + 1). Let the frequency of discordant pairs in which the case is exposed be f_{10} and the frequency of discordant pairs in which the control is exposed be f_{01}. Since a discordant pair must contribute either to f_{10} or to f_{01}, we can treat the distribution of discordant pairs of type B as binomial; the likelihood of observing exactly f_{10} type B pairs, given that there are $f_{10} + f_{01}$ discordant pairs is then

$$\Pr = \binom{f_{10} + f_{01}}{f_{10}} \left(\frac{OR}{OR + 1}\right)^{f_{10}} \left(\frac{1}{OR + 1}\right)^{f_{01}} \qquad [13\text{-}1]$$

The maximum likelihood estimator of the OR is derived by maximizing the above expression with respect to the OR. The maximization is equivalent to maximizing the logarithm of expression 13-1,

$$\ln(\Pr) = \ln\binom{f_{10} + f_{01}}{f_{10}} + f_{10} \ln\left(\frac{OR}{OR + 1}\right) + f_{01} \ln\left(\frac{1}{OR + 1}\right)$$

Taking the derivative and setting it equal to zero gives

$$\frac{d(\ln(\Pr))}{d(OR)} = 0 = f_{10} - (\widehat{OR})f_{01} \qquad [13\text{-}2]$$

$$\widehat{OR} = \frac{f_{10}}{f_{01}}$$

which is the conditional maximum likelihood estimator.

Alternatively, the Mantel-Haenszel estimator of the odds ratio can be used (formula 12-26). For each table of type B, $a_i d_i/T_i = \frac{1}{2}$ and $b_i c_i/T_i = 0$. For each table of type C, $a_i d_i/T_i = 0$ and $b_i c_i/T_i = \frac{1}{2}$. Therefore the Mantel-Haenszel estimator for matched-pair data is

$$\widehat{OR}_{MH} = \frac{\Sigma a_i d_i/T_i}{\Sigma b_i c_i/T_i} = \frac{\frac{1}{2} f_{10}}{\frac{1}{2} f_{01}} = \frac{f_{10}}{f_{01}}$$

the same expression as the maximum likelihood estimator.

POINT ESTIMATION OF RELATIVE RISK WITH R CONTROLS MATCHED TO EACH CASE

For the more general situation of R controls matched to each case, there is a larger number of possible exposure patterns, the exact number depending on the value of R. Considering all R controls as equivalent, there are R + 1 different outcomes possible for each matched set of controls, corresponding to the number of controls in the matched set that are exposed and ranging from zero exposed at one extreme to R exposed at the other extreme. Since the case can be either exposed or unexposed, the total number of possible exposure patterns is 2 (R + 1). A convenient way to summarize the data is simply to tally the frequency of matched sets with each exposure pattern, using the notation of Table 13-6.

The frequency f_{00} is the number of matched sets with no exposed subjects: these elementary strata have a zero marginal total and do not contribute to the analysis. Similarly, f_{1R} refers to the sets with no unexposed subjects; these sets also have a zero marginal total and do not contribute to the analysis. The remaining 2R types of sets are all informative sets, representing elementary 2 × 2 tables with nonzero marginal totals. Note that as R increases, the probability that a given set will be informative also increases, since the likelihood that all the controls will have the same exposure as the case becomes smaller. If there is a 90 percent probability that a matched control has an exposure history concordant with that of the case, the probability that the matched set for that case will contribute to the analysis ranges from 10 percent for R = 1 to $1 - (0.9)^5 = 41$ percent for R = 5. If a matched control has an 80 percent probability of having a concordant exposure, the probability that a set is informative ranges from 20 percent for R = 1 to 67 percent for R = 5.

Let us denote the total number of exposed subjects in a matched set as m. The value of m ranges from zero to R + 1, but the informative sets are

Table 13-6. Data summary for R controls matched to each case, indicating the frequency (f_{ij}) of matched sets with every possible exposure pattern

	No. exposed controls					
	0	1	2	3	...	R
Exposed cases	f_{10}	f_{11}	f_{12}	f_{13}	...	f_{1R}
Unexposed cases	f_{00}	f_{01}	f_{02}	f_{03}	...	f_{0R}

those for which $1 \leq m \leq R$, keeping all marginal totals nonzero. For a given value of m, there are two possible patterns of exposure for the matched set, corresponding to the case being exposed and $m - 1$ controls being exposed, or the case not being exposed and m controls being exposed. From the noncentral hypergeometric distribution, the probability that the case is exposed, given m exposed subjects, is

$$\Pr(\text{case is exposed, given m}) = \frac{m(OR)}{R + 1 - m + m(OR)} = \frac{OR}{\frac{R + 1 - m}{m} + OR} \qquad [13\text{-}3]$$

and the probability that the case is unexposed is the complement,

$$\Pr(\text{case is unexposed, given m}) = \frac{(R + 1 - m)/m}{\frac{R + 1 - m}{m} + OR} \qquad [13\text{-}4]$$

It is again convenient to consider the observations as following a binomial distribution, but with R-to-1 matching there is a separate binomial distribution for each value of m. Thus, for $m = 1$, there is a total of $f_{10} + f_{01}$ sets, and, given that exactly one subject in a set is exposed, the probability of exposure is OR/(OR + R), from equation 13-3. The probability of observing exactly f_{10} and f_{01} sets with the case exposed and unexposed, respectively, given a total of $f_{10} + f_{01}$ sets with one exposed subject, is

$$\Pr = \binom{f_{10} + f_{01}}{f_{10}} \left(\frac{OR}{OR + R}\right)^{f_{10}} \left(\frac{R}{OR + R}\right)^{f_{01}}$$

The overall likelihood for the data is the product of the binomial probabilities corresponding to each value of m from 1 to R:

$$\Pr = \prod_{m=1}^{R} \binom{f_{1,m-1} + f_{0,m}}{f_{1,m-1}} \left(\frac{OR}{(R + 1 - m)/m + OR}\right)^{f_{1,m-1}} \cdot \left(\frac{(R + 1 - m)/m}{(R + 1 - m)/m + OR}\right)^{f_{0,m}} \qquad [13\text{-}5]$$

The logarithm of the above likelihood expression is

$$\ln(\Pr) = \sum_{m=1}^{R} \left[\ln\binom{f_{1,m-1} + f_{0,m}}{f_{1,m-1}} + f_{1,m-1} \ln\left(\frac{OR}{(R + 1 - m)/m + OR}\right) \right.$$

$$\left. + f_{0,m} \ln\left(\frac{(R + 1 - m)/m}{(R + 1 - m)/m + OR}\right) \right]$$

Taking the derivative of the above expression with regard to the OR and setting it equal to zero yields the equation for the conditional maximum likelihood solution for the OR [Miettinen, 1970]:

$$\frac{\sum_{m=1}^{R} f_{1,m-1}}{\widehat{OR}} - \sum_{m=1}^{R} \frac{f_{1,m-1} + f_{0,m}}{(R + 1 - m)/m + \widehat{OR}} = 0 \qquad [13\text{-}6]$$

Equation 13-6 reduces to equation 13-2 for $R = 1$. For $R = 2$, it can be solved explicitly for $\widehat{OR}$ [Miettinen, 1970], but for values of R greater than 2 an iterative solution is necessary. Even so, equation 13-6 represents a rather simple computational exercise compared with the onerous computations needed to obtain a conditional maximum likelihood estimate of the odds ratio for unmatched stratified data.

The data in Example 13-1 represent the individual exposure values for each subject in a matched case-control study with 18 cases and four controls matched to each case. The cases were women with ectopic pregnancy; controls were women without ectopic pregnancy drawn from the same source population and matched individually to the cases for number of pregnancies, age, and husband's level of education. All subjects had had at least one previous pregnancy. A positive history indicates that the woman had at least one induced abortion.

If only the first control had been matched to each case, the investigators would have observed nine concordant pairs (four concordant pairs with positive exposure histories and five with negative exposure histories) and nine discordant pairs. In eight of the discordant pairs the case is exposed, compared with only one in which the control is exposed, giving a relative risk estimate of $8/1 = 8$. Considering all controls that were studied, there are $2(4 + 1) = 10$ types of exposure patterns for the matched sets. The distribution of exposure patterns for the data in Example 13-1 is shown in Table 13-7.

Six of the matched sets have completely concordant exposure histories and so are noncontributory to the analysis. The data from the remaining sets can be used to estimate the odds ratio, or relative risk, of ectopic pregnancy after induced abortion by substituting into equation 13-6:

$$\frac{11}{\widehat{OR}} - \frac{4}{\widehat{OR} + 4} - \frac{5}{\widehat{OR} + 3/2} - \frac{3}{\widehat{OR} + 2/3} = 0$$

A trial and error solution gives $\widehat{OR} = 23$.

Example 13-1. Previous history of induced abortion among women with ectopic pregnancy and matched controls. Data of Trichopoulos et al. [Miettinen, 1969]

	Control			
Case	1	2	3	4
−	−	−	−	−
+	−	+	−	−
+	−	−	−	−
−	−	−	−	−
−	+	−	−	−
+	−	−	−	−
+	−	−	−	−
−	−	−	−	−
+	+	−	−	+
+	−	+	−	−
+	−	+	+	−
−	−	−	−	−
+	+	+	+	+
+	−	−	+	−
+	−	−	+	−
+	+	−	−	−
−	−	−	−	−
+	+	−	−	+

\+ = previous induced abortion; − = no previous induced abortion.

Table 13-7. Pattern of exposure for the 18 matched sets in example 13-1

	No. exposed controls				
	0	1	2	3	4
Exposed cases	3	5	3	0	1
Unexposed cases	5	1	0	0	0

An alternative to the maximum likelihood approach to estimation is the Mantel-Haenszel approach. When the matching ratio exceeds one control per case, the two approaches are not identical. With R-to-1 matching, formula 12-26 can be rewritten as follows:

$$\hat{OR}_{MH} = \frac{\sum_{m=1}^{R} (R + 1 - m) f_{1,m-1}}{\sum_{m=1}^{R} m f_{0,m}} \qquad [13\text{-}7]$$

Applying formula 13-7 to the data in Table 13-7 gives

$$\hat{OR}_{MH} = \frac{4(3) + 3(5) + 2(3)}{1(1)} = \frac{33}{1} = 33$$

which differs noticeably from the conditional maximum likelihood estimate of 23. The large difference between the two estimates is attributable to the fact that there are only 12 informative sets, and 11 of these are supportive of a positive association, representing an extreme result with somewhat scanty data. Consequently, it is not surprising that two different estimators give somewhat discrepant results. Breslow [1981] has shown that statistically the Mantel-Haenszel estimator is consistent for matched data, is as efficient as the conditional maximum likelihood approach when the OR = 1, and is nearly as efficient over a wide range of conditions.

POINT ESTIMATION OF RELATIVE RISK WITH A VARYING NUMBER OF CONTROLS MATCHED TO EACH CASE

With a varying number of controls matched to each case, the data can be summarized by a set of displays like the one in Table 13-6, each one corresponding to a different value of R. The likelihood for the data is the product of the likelihood expressions corresponding to each value of R, and the equation that yields the maximum likelihood estimate of the OR is a simple extension of equation 13-6:

$$\sum_{R} \left[\frac{\sum_{m=1}^{R} f_{1,m-1}}{\hat{OR}} - \sum_{m=1}^{R} \frac{f_{1,m-1} + f_{0,m}}{(R + 1 - m)/m + \hat{OR}} \right] = 0 \qquad [13\text{-}8]$$

The data in Example 13-2 are derived from a study of myocardial infarction and history of coffee consumption [Jick et al., 1973]. The authors attempted to match two controls to each case, but for 27 cases only one

Example 13-2. Distribution of cases of myocardial infarction and matched controls according to amount of coffee drinking; subjects drinking one to five cups of coffee per day were excluded [Jick et al., 1973]

	No. controls drinking 6+ cups/day				
	Matched pairs		Matched triplets		
	0	1	0	1	2
Cases					
6+ cups/day	8	8	16	23	4
0 cups/day	8	3	20	22	3

matched control was available. The resulting data consist of 27 matched pairs and 88 matched triplets. The use of these data and equation 13-8 to determine the maximum likelihood estimate of the odds ratio produces the following likelihood equation:

$$\left[\frac{8}{\widehat{OR}} - \frac{11}{\widehat{OR} + 1}\right] + \left[\frac{39}{\widehat{OR}} - \frac{38}{\widehat{OR} + 2} - \frac{26}{\widehat{OR} + 1/2}\right] = 0$$

Solving the above equation by trial and error gives $\widehat{OR} = 2.0$.

The generalization of formula 13-7 for the Mantel-Haenszel estimator of the odds ratio with matched case-control data having a varying number of controls, R, matched to each case can be derived easily from formula 12-26:

$$\widehat{OR}_{MH} = \frac{\sum_R \sum_{m=1}^{R} \frac{(R + 1 - m)f_{1,m-1}}{R + 1}}{\sum_R \sum_{m=1}^{R} \frac{mf_{0,m}}{R + 1}} \qquad [13\text{-}9]$$

Applying the above formula to the data of Example 13-2 gives the Mantel-Haenszel estimate as

$$\widehat{OR}_{MH} = \frac{8/2 + [2(16) + 23]/3}{3/2 + [22 + 2(3)]/3} = 2.1$$

which is nearly identical to the maximum likelihood estimate and is considerably easier to obtain.

Statistical Hypothesis Testing with Matched Case-Control Data

Since the analysis of matched case-control data is equivalent to an analysis stratifying the data according to the matching factors, hypothesis testing for matched data is accomplished simply by applying the general approach for stratified data to the strata defined by the matching. As with point estimation, some of the formulas can be simplified for matching because the 2×2 tables can have only a limited number of configurations; since a matched analysis typically involves many strata, the simplifications may prove important. Even the two tableaus for displaying the data in Example 13-2 illustrate this point because they summarize data on 115 strata, corresponding to the 115 matched sets.

HYPOTHESIS TESTING FOR MATCHED CASE-CONTROL PAIRS

For matched pairs, an exact *P*-value can be calculated from equation 13-1 by setting OR = 1 and calculating the tail probability. This calculation is simply the tail probability of a binomial distribution with a probability of 0.5 for each binomial trial. The tail probability for the Fisher *P*-value is

$$\text{Fisher } P = \sum_{k=f_{10}}^{f_{10}+f_{01}} \binom{f_{10} + f_{01}}{k} \left(\frac{1}{2}\right)^{f_{10}+f_{01}} \qquad [13\text{-}10]$$

for $f_{10} \geq f_{01}$. If $f_{01} > f_{10}$, then the lower tail should be calculated by summing over the range $0 \leq k \leq f_{10}$.

To get the exact mid-*P* value, only half the probability of the observed data should be included. For the upper tail, this modification gives

$$\text{Mid-}P = \frac{1}{2}\binom{f_{10} + f_{01}}{f_{10}}\left(\frac{1}{2}\right)^{f_{10}+f_{01}} + \sum_{k=f_{10}+1}^{f_{10}+f_{01}} \binom{f_{10} + f_{01}}{k} \left(\frac{1}{2}\right)^{f_{10}+f_{01}} \qquad [13\text{-}11]$$

If $f_{01} > f_{10}$, then the lower tail should be calculated by summing over the range $0 \leq k \leq f_{10} - 1$.

Consider the data in Example 13-2 relating just to matched pairs. The 16 pairs for which the exposure history was concordant do not contribute to the evaluation and should be ignored. Of the remaining 11 pairs, 8 are discordant with the case exposed. The exact Fisher one-tail *P*-value is, from formula 13-10,

$$\binom{11}{8}\left(\frac{1}{2}\right)^{11} + \binom{11}{9}\left(\frac{1}{2}\right)^{11} + \binom{11}{10}\left(\frac{1}{2}\right)^{11} + \binom{11}{11}\left(\frac{1}{2}\right)^{11}$$

$$= 0.0806 + 0.0269 + 0.0054 + 0.0005 = 0.11$$

The mid-*P* value is the same summation except for the first term, which would be ½(0.0806), giving a one-tail *P*-value of 0.07.

An approximate *P*-value can be calculated using the Mantel-Haenszel test statistic (formula 12-38). For matched pairs, the Mantel-Haenszel test simplifies to

$$\chi = \frac{f_{10} - f_{01}}{\sqrt{f_{10} + f_{01}}} \qquad [13\text{-}12]$$

which is a form of the test first described by McNemar [1947] and often referred to as the McNemar test.

For the matched pair data in Example 13-2, this test formula gives

$$\chi = \frac{8 - 3}{\sqrt{11}} = 1.51$$

which corresponds to a one-tail *P*-value of 0.07, agreeing well with the exact mid-*P* value even for these apparently small numbers.

HYPOTHESIS TESTING FOR R CONTROLS MATCHED TO EACH CASE

Exact hypothesis testing for R controls matched to each case is considerably more complicated than hypothesis testing for matched pairs. The data can be considered a set of R binomial distributions with the likelihood function expressed in formula 13-5. For hypothesis testing, the value of the odds ratio in expression 13-5 is set equal to unity. The upper tail probability is determined by evaluating formula 13-5 for every possible distribution of the data for which the number of exposed cases is equal to or greater than the number observed (the exposed cases for whom all matched controls are also exposed can be ignored). The exact Fisher *P*-value is therefore

Fisher upper-tail probability

$$= \sum_{k=a}^{M_1} \prod_{m=1}^{R} \binom{f_{1,m-1} + f_{0,m}}{k_m} \left(\frac{m}{R+1}\right)^{k_m} \left(\frac{R+1-m}{R+1}\right)^{f_{1,m-1}+f_{0,m}-k_m} \qquad [13\text{-}13]$$

where a is the total number of exposed cases in sets with at least one unexposed control, M_1 is the total number of sets that are not completely concordant, k_m is the permutation of the possible number of exposed cases with $m-1$ exposed controls, i.e., the total number of exposed cases that could have been observed among matched sets that actually had m exposed subjects, and k is the sum of k_m:

$$a = \sum_{m=1}^{R} f_{1,m-1}$$

$$M_1 = \sum_{m=1}^{R} f_{1,m-1} + f_{0,m}$$

$$k = \sum_{m=1}^{R} k_m$$

The tail summation includes all the combinations of the data that could give rise to all the values of k in the range from a to M_1. For the lower tail probability, the range of summation for k is from 0 to a.

To obtain the exact mid-*P* value, it is necessary to include only half the probability for $k = a$, as follows:

Mid-*P* upper tail probability

$$= \frac{1}{2} \sum_{k=a} \prod_{m=1}^{R} \binom{f_{1,m-1} + f_{0,m}}{k_m} \left(\frac{m}{R+1}\right)^{k_m} \left(\frac{R+1-m}{R+1}\right)^{f_{1,m-1}+f_{0,m}-k_m}$$
$$+ \sum_{k=a+1}^{M_1} \prod_{m=1}^{R} \binom{f_{1,m-1} + f_{0,m}}{k_m} \left(\frac{m}{R+1}\right)^{k_m} \left(\frac{R+1-m}{R+1}\right)^{f_{1,m-1}+f_{0,m}-k_m} \qquad [13\text{-}14]$$

For the lower tail mid-*P* value, the second summation in equation 13-14 should be for $0 \leq k \leq a-1$ rather than $a+1 \leq k \leq M_1$.

Consider the data in Example 13-1 (Table 13-7). Disregarding the exposed case that had four exposed matched controls, there are 11 exposed cases, so $a = 11$. The total number of informative sets, M_1, is $3 + 5 + 3 + 0 + 1 + 0 + 0 + 0 = 12$. There is only one set of values for $\{k_m\}$ for which every informative set has an exposed case; since there are four, five, three, and zero sets, respectively, for $m = 1, 2, 3$, and 4, the values of k_m would be 4, 5, 3, and 0, respectively, to obtain the most extreme outcome, with all the cases exposed. There are three different patterns that could yield $\Sigma k_m = 11$. These are, starting with the observed pattern, 3, 5, 3, 0; 4, 4, 3, 0; and 4, 5, 2, 0. Thus the upper tail has four possible outcomes in it, including the observed data; there are two equally extreme outcomes, and one more extreme outcome.

Let us calculate the probability of each of these four outcomes. Consider first the most extreme outcome, 4, 5, 3, 0. The probability is, using expression 13-13,

$$\binom{4}{4}\left(\frac{1}{5}\right)^4\left(\frac{4}{5}\right)^0 \cdot \binom{5}{5}\left(\frac{2}{5}\right)^5\left(\frac{3}{5}\right)^0 \cdot \binom{3}{3}\left(\frac{3}{5}\right)^3\left(\frac{2}{5}\right)^0$$
$$= \left(\frac{1}{5}\right)^4\left(\frac{2}{5}\right)^5\left(\frac{3}{5}\right)^3 = 0.00000354$$

For the observed data, the probability is

$$\binom{4}{3}\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^1 \cdot \binom{5}{5}\left(\frac{2}{5}\right)^5\left(\frac{3}{5}\right)^0 \cdot \binom{3}{3}\left(\frac{3}{5}\right)^3\left(\frac{2}{5}\right)^0$$
$$= 4 \cdot 4\left(\frac{1}{5}\right)^4\left(\frac{2}{5}\right)^5\left(\frac{3}{5}\right)^3 = 0.00005662$$

For the remaining two possible outcomes, the probabilities are

$$\binom{4}{4}\left(\frac{1}{5}\right)^4\left(\frac{4}{5}\right)^0 \cdot \binom{5}{4}\left(\frac{2}{5}\right)^4\left(\frac{3}{5}\right)^1 \cdot \binom{3}{3}\left(\frac{3}{5}\right)^3\left(\frac{2}{5}\right)^0$$
$$= 5\left(\frac{1}{5}\right)^4\left(\frac{2}{5}\right)^4\left(\frac{3}{5}\right)^4 = 0.00002654$$

and

$$\binom{4}{4}\left(\frac{1}{5}\right)^4\left(\frac{4}{5}\right)^0 \cdot \binom{5}{5}\left(\frac{2}{5}\right)^5\left(\frac{3}{5}\right)^0 \cdot \binom{3}{2}\left(\frac{3}{5}\right)^2\left(\frac{2}{5}\right)^1$$
$$= 3\left(\frac{1}{5}\right)^4\left(\frac{2}{5}\right)^6\left(\frac{3}{5}\right)^2 = 0.00000708$$

The sum of the probabilities for the four possible outcomes is 0.000094, which is the upper Fisher exact *P*-value. The upper mid-*P* would include just half the probability for the outcomes for which $\Sigma k_m = 11$, namely,

$$\frac{1}{2}(0.00005662 + 0.00002654 + 0.00000708) + 0.00000354 = 0.000049$$

which is barely more than half of the Fisher exact *P*-value, since the observed outcome is the second most extreme possible outcome in the upper tail.

Approximate hypothesis testing for R controls matched to each case can once again be performed with the Mantel-Haenszel test (formula 12-38). With one case and R controls in each stratum, the test can be rewritten in terms of the distribution of matched sets as

$$\chi = \frac{\sum_{m=1}^{R}\left[f_{1,m-1} - \frac{m}{R+1}(f_{1,m-1} + f_{0,m})\right]}{\sqrt{\sum_{m=1}^{R}(f_{1,m-1} + f_{0,m})\frac{m(R+1-m)}{(R+1)^2}}} \qquad [13\text{-}15]$$

For the data in Table 13-7, the above formula yields

$$\chi = \frac{11 - \frac{23}{5}}{\sqrt{\frac{16}{25} + \frac{30}{25} + \frac{18}{25}}} = 4.0$$

which corresponds to a one-tail *P*-value of 0.00003. This value agrees reasonably well with the exact upper mid-*P* value considering the small numbers involved and the extremeness of the observed outcome.

HYPOTHESIS TESTING WITH A VARYING NUMBER OF CONTROLS MATCHED TO EACH CASE

Formulas 13-13 and 13-14 can be easily extended to accommodate a varying ratio of controls to cases by taking the product of the binomial probabilities over all values for R:

Fisher upper-tail probability

$$= \sum_{k=a}^{M_1}\prod_{R}\prod_{m=1}^{R}\binom{f_{1,m-1} + f_{0,m}}{k_m}\left(\frac{m}{R+1}\right)^{k_m}\left(\frac{R+1-m}{R+1}\right)^{f_{1,m-1}+f_{0,m}-k_m} \qquad [13\text{-}16]$$

Mid-*P* upper-tail probability

$$= \frac{1}{2}\sum_{k=a}^{M_1}\prod_{R}\prod_{m=1}^{R}\binom{f_{1,m-1} + f_{0,m}}{k_m}\left(\frac{m}{R+1}\right)^{k_m}\left(\frac{R+1-m}{R+1}\right)^{f_{1,m-1}+f_{0,m}-k_m} + \sum_{k=a+1}^{M_1}\prod_{R}\prod_{m=1}^{R}\binom{f_{1,m-1} + f_{0,m}}{k_m}\left(\frac{m}{R+1}\right)^{k_m}\left(\frac{R+1-m}{R+1}\right)^{f_{1,m-1}+f_{0,m}-k_m} \qquad [13\text{-}17]$$

where the notation is identical to that of formulas 13-13 and 13-14. Although formulas 13-16 and 13-17 are similar to formulas 13-13 and 13-14, their application is considerably more difficult because the number of possible outcomes corresponding to each value of k can be high. In Example 13-2, there is one binomial for R = 1 and two binomials for R = 2 to consider in the calculations; the total number of possible outcomes for the three binomials is 11 × 38 × 26 = 10,868. Of these, 3522 are in the upper-tail summation for formulas 13-16 and 13-17. There are 279 combinations of the three binomials that yield exactly 47 exposed cases, the observed number. Taking the sum of 3522 terms, each of which is the product of three binomial probabilities, is a task for a computer. An evaluation of expression 13-16 for the data in Example 13-2 gives 0.0038 for the upper-tail Fisher *P*-value and 0.0028 for the upper mid-*P* value.

The approximate test statistic in formula 13-15 can also be easily extended to accommodate a varying number of controls by taking the sums in the numerator and denominator over the values of R:

$$\chi = \frac{\sum_{R}\sum_{m=1}^{R}\left[f_{1,m-1} - \frac{m}{R+1}(f_{1,m-1} + f_{0,m})\right]}{\sqrt{\sum_{R}\sum_{m=1}^{R}(f_{1,m-1} + f_{0,m})\frac{m(R+1-m)}{(R+1)^2}}} \qquad [13\text{-}18]$$

For example 13-2, the approximate formula above gives

$$\chi = \frac{8 - \left(\frac{1}{2}\right)(11) + 16 - \left(\frac{1}{3}\right)(38) + 23 - \left(\frac{2}{3}\right)(26)}{\sqrt{11\left(\frac{1}{4}\right) + 38\left(\frac{2}{9}\right) + 26\left(\frac{2}{9}\right)}} = 2.79$$

which corresponds to a one-tail *P*-value of 0.0026, in close agreement with the upper mid-*P* value.

val Estimation of the Odds Ratio with Matched Case-Control Data

INTERVAL ESTIMATION FOR MATCHED CASE-CONTROL PAIRS

Exact confidence limits for the odds ratio from matched case-control pairs can be calculated based on the probability distribution of the possible discordant pairs, conditional on the total number of discordant pairs, by expressing the probability as a function of the odds ratio (see formula 13-1):

$$\alpha/2 = \sum_{k=f_{10}}^{f_{10}+f_{01}} \binom{f_{10}+f_{01}}{k} \left(\frac{\underline{OR}}{\underline{OR}+1}\right)^{k} \left(\frac{1}{\underline{OR}+1}\right)^{f_{10}+f_{01}-k} \qquad [13\text{-}19]$$

$$1-\alpha/2 = \sum_{k=f_{10}+1}^{f_{10}+f_{01}} \binom{f_{10}+f_{01}}{k} \left(\frac{\overline{OR}}{\overline{OR}+1}\right)^{k} \left(\frac{1}{\overline{OR}+1}\right)^{f_{10}+f_{01}-k} \qquad [13\text{-}20]$$

The above formulas, when solved for $\underline{OR}$ and $\overline{OR}$, give the exact Fisher confidence limits. To obtain the mid-*P* exact confidence limits, only half the probability that $k = f_{10}$ is added to the tail:

$$\alpha/2 = \frac{1}{2}\binom{f_{10}+f_{01}}{f_{10}} \left(\frac{\underline{OR}}{\underline{OR}+1}\right)^{f_{10}} \left(\frac{1}{\underline{OR}+1}\right)^{f_{01}} + \sum_{k=f_{10}+1}^{f_{10}+f_{01}} \binom{f_{10}+f_{01}}{k} \left(\frac{\underline{OR}}{\underline{OR}+1}\right)^{k} \left(\frac{1}{\underline{OR}+1}\right)^{f_{10}+f_{01}-k} \qquad [13\text{-}21]$$

and

$$1-\alpha/2 = \frac{1}{2}\binom{f_{10}+f_{01}}{f_{10}} \left(\frac{\overline{OR}}{\overline{OR}+1}\right)^{f_{10}} \left(\frac{1}{\overline{OR}+1}\right)^{f_{01}} + \sum_{k=f_{10}+1}^{f_{10}+f_{01}} \binom{f_{10}+f_{01}}{k} \left(\frac{\overline{OR}}{\overline{OR}+1}\right)^{k} \left(\frac{1}{\overline{OR}+1}\right)^{f_{10}+f_{01}-k} \qquad [13\text{-}22]$$

The solution of equations 13-19 and 13-20 or 13-21 and 13-22 amounts to finding the exact confidence limits for a binomial parameter, p, which is a function of the odds ratio: $p = OR/(OR + 1)$. Consider the data in Example 13-2 relating to matched pairs. With 11 discordant pairs, 8 of which have an exposed case, the calculation of exact confidence limits for the odds ratio corresponds to setting exact confidence limits for the binomial parameter estimated by eight successes in 11 trials. The Fisher exact 90 percent confidence limits are, from formulas 13-19 and 13-20, 0.4356 and 0.9212 for the binomial parameter, which correspond to a 90 percent exact Fisher confidence interval of 0.77 and 11.7 for the odds ratio. If formulas 13-21 and 13-22 are used to get the mid-*P* exact limits, the results are 0.4702 and 0.9030 for the binomial parameter, corresponding

to 0.89 and 9.31 for the 90 percent exact limits for the odds ratio. The wide limits reflect the small number of discordant pairs.

Approximate confidence limits for matched case-control pairs can be determined in several ways. One approach is to determine the confidence limits for the probability that a discordant pair has an exposed case, based on the large sample characteristics of the binomial distribution, and then convert these confidence limits to the corresponding limits for the odds ratio. Other approaches include the large sample characteristics of maximum likelihood estimators, the formula by Robins et al. [1986] for the variance of the logarithm of the Mantel-Haenszel estimate (formula 12-58), and the test-based procedure.

First let us consider basing the approximation on the sampling distribution of the binomial distribution, which has a variance of pq/n for large n, where n is the number of binomial trials, p is the probability of a "success," and $q = 1 - p$. For matched case-control pairs, $\hat{p} = f_{10}/(f_{10} + f_{01})$, and confidence limits for $\hat{p}$ can be approximated by

$$\frac{f_{10}}{f_{10}+f_{01}} \pm Z\sqrt{\frac{f_{10}f_{01}}{(f_{10}+f_{01})^3}} \qquad [13\text{-}23]$$

where Z is the value of the standard normal distribution corresponding to the desired level of confidence, the plus sign gives the upper confidence limit, and the minus sign gives the lower confidence limit. The corresponding limits for the odds ratio are given by $\underline{OR} = \underline{p}/(1 - \underline{p})$ and $\overline{OR} = \overline{p}/(1 - \overline{p})$, or

$$\underline{OR} = \frac{\dfrac{f_{10}}{f_{10}+f_{01}} - Z\sqrt{\dfrac{f_{10}f_{01}}{(f_{10}+f_{01})^3}}}{1 - \dfrac{f_{10}}{f_{10}+f_{01}} + Z\sqrt{\dfrac{f_{10}f_{01}}{(f_{10}+f_{01})^3}}} \qquad [13\text{-}24]$$

$$\overline{OR} = \frac{\dfrac{f_{10}}{f_{10}+f_{01}} + Z\sqrt{\dfrac{f_{10}f_{01}}{(f_{10}+f_{01})^3}}}{1 - \dfrac{f_{10}}{f_{10}+f_{01}} - Z\sqrt{\dfrac{f_{10}f_{01}}{(f_{10}+f_{01})^3}}} \qquad [13\text{-}25]$$

The above approximate confidence limits are simple to calculate, but they are inaccurate unless the number of discordant pairs is reasonably large. For values of the odds ratio that are far from the null value, the number of discordant pairs must be very large for the approximation to be adequate. The difficulty is that the binomial distribution does not approximate

a normal distribution very well if the number of trials is modest, especially if the probability of a success is far from 0.5. Formula 13-23 always produces confidence limits for p that are symmetric about $\hat{p}$ despite the fact that the sampling distribution can be strikingly asymmetric for values of p that depart from 0.5, the center of the range of the distribution. It is possible to calculate a confidence interval from formula 13-23 with a boundary outside the admissible range of 0 to 1 for p. For example, if two successes were observed in 10 trials, formula 13-23 gives a 90 percent confidence interval for $\hat{p}$ with a lower bound of -0.008; eight successes in 10 trials would give, from the same formula, an upper bound of 1.008. These limits outside the admissible range for p correspond to negative values of the odds ratio as determined from formulas 13-24 and 13-25.

A more accurate method for obtaining approximate confidence limits for the binomial parameter was proposed by Wilson [1927]. This approach takes into account the asymmetry of the distribution and consequently never gives results outside the admissible range. Wilson's formula is

$$\frac{T}{T + Z^2}\left[\frac{f_{10}}{T} + \frac{Z^2}{2T} \pm Z\sqrt{\frac{f_{10}f_{01}}{T^3} + \frac{Z^2}{4T^2}}\right] \qquad [13\text{-}26]$$

where T is $f_{10}+f_{01}$, Z is $Z_{1-\alpha/2}$, the plus sign gives the upper confidence limit for p, and the minus sign gives the lower confidence limit for p. Confidence limits for the odds ratio are taken, as before, as $\underline{p}/(1-\underline{p})$ and $\overline{p}/(1-\overline{p})$. If $f_{10} = 8$ and $f_{01} = 2$, the 90 percent confidence limits for p from formula 13-26 are 0.541 and 0.931, well within the admissible range and reflective of the asymmetry of the sampling distribution.

For the 11 discordant matched pairs in the data of Example 13-2, formula 13-23 gives the 90 percent confidence limits of the binomial parameter of 0.506 and 0.948, corresponding to 1.03 and 18.3 for the odds ratio. These limits, especially the upper one, agree poorly with the exact limits calculated earlier. Formula 13-26, on the other hand, gives a 90 percent confidence interval for the binomial parameter of 0.479 and 0.885, corresponding to a confidence interval for the odds ratio of 0.92 to 7.72, which agrees much more closely with the mid-*P* exact 90 percent interval.

The ratio of discordant matched pairs is simultaneously the maximum likelihood estimate and the Mantel-Haenszel estimate of the odds ratio. For matched case-control data the variance of the maximum likelihood estimate of the odds ratio has been described by Miettinen [1970]. As is usual for ratio estimators, the confidence limits are set for the logarithmic transformation of the estimate, and then the transformation is reversed. For matched pairs, the large sample formula for the variance of the logarithm of the odds ratio is

$$\text{Var}[\ln(\widehat{OR})] \doteq \frac{f_{10} + f_{01}}{f_{10}f_{01}} \qquad [13\text{-}27]$$

which gives approximate confidence limits for the odds ratio of

$$\underline{OR} = \exp\left[\ln\left(\frac{f_{10}}{f_{01}}\right) - Z\sqrt{\frac{f_{10} + f_{01}}{f_{10}f_{01}}}\right] \qquad [13\text{-}28]$$

and

$$\overline{OR} = \exp\left[\ln\left(\frac{f_{10}}{f_{01}}\right) + Z\sqrt{\frac{f_{10} + f_{01}}{f_{10}f_{01}}}\right] \qquad [13\text{-}29]$$

For the matched pair data in Example 13-2, the variance is estimated from formula 13-27 to be $11/24 = 0.458$, and the 90 percent confidence limits from formulas 13-28 and 13-29 are 0.88 and 8.12. Considering the few pairs involved, this approximation gives excellent results for these data; the lower bound is nearly equal to the mid-*P* exact lower limit, and the upper bound is reasonably close to the corresponding exact upper limit.

Since the maximum likelihood and Mantel-Haenszel estimators are the same for matched case-control pairs, it is not surprising to find that formula 12-58 for the variance of the logarithm of the Mantel-Haenszel estimator is identical to formula 13-27 when applied to matched pairs.

One other approach to approximate confidence limits for the odds ratio estimated from matched case-control pairs is the test-based approach. For matched pairs, the test-based limits are

$$\left(\frac{f_{10}}{f_{01}}\right)^{(1 \pm Z/\chi)} \qquad [13\text{-}30]$$

where the χ is the value from equation 13-12. Since equations 13-28 and 13-29 represent a straightforward and theoretically optimal approach to obtaining approximate confidence limits for matched case-control pairs, it is generally preferable to use them rather than the test-based approach, the binomial formulations in equations 13-24 through 13-26, or other alternatives. For comparison, the 90 percent test-based confidence limits for the matched pair data in Example 13-2 are 0.91 and 7.78, which are similar to the results obtained from formula 13-26 and slightly worse, compared with the mid-*P* exact limits, than the results obtained from equations 13-28 and 13-29.

INTERVAL ESTIMATION FOR R CONTROLS MATCHED TO EACH CASE

Exact interval estimation of the odds ratio with R matched controls for each case proceeds from the probability expression for the data written as a function of the odds ratio:

$$\Pr(\text{data}) = \prod_{m=1}^{R} \binom{f_{1,m-1} + f_{0,m}}{f_{1,m-1}} \left(\frac{OR}{OR + C_m}\right)^{f_{1,m-1}} \left(\frac{C_m}{OR + C_m}\right)^{f_{0,m}} \quad [13\text{-}31]$$

where $C_m = (R + 1 - m)/m$ and the remaining notation follows Table 13-6. Expression 13-31 represents the product of R binomial probabilities, in which the binomial parameter corresponding to the probability of a "success" (i.e., a matched set with an exposed case, given that the set has m exposed subjects) is

$$\Pr(\text{exposed case given m exposed subjects}) = \frac{OR}{OR + (R + 1 - m)/m}$$

which can be derived from the noncentral hypergeometric distribution. The exact confidence limits are determined iteratively by summing the value of expression 13-31 for every possible outcome of the data that departs equally or more extremely from the null hypothesis, starting with the observed data; the sum is calculated for trial values of the odds ratio until the tail area equals the desired value. Thus, the Fisher exact limits are the solutions to the following equations:

$$\alpha/2 = \sum_{k=a}^{M_1} \prod_{m=1}^{R} \binom{f_{1,m-1} + f_{0,m}}{k_m} \left(\frac{OR}{OR + C_m}\right)^{k_m} \left(\frac{C_m}{OR + C_m}\right)^{f_{1,m-1} + f_{0,m} - k_m} \quad [13\text{-}32]$$

and

$$1 - \alpha/2 = \sum_{k=a+1}^{M_1} \prod_{m=1}^{R} \binom{f_{1,m-1} + f_{0,m}}{k_m} \left(\frac{\overline{OR}}{\overline{OR} + C_m}\right)^{k_m} \left(\frac{C_m}{\overline{OR} + C_m}\right)^{f_{1,m-1} + f_{0,m} - k_m} \quad [13\text{-}33]$$

where a is the total number of exposed cases in sets with at least one unexposed control,

$$a = \sum_{m=1}^{R} f_{1,m-1} \qquad M_1 = \sum_{m=1}^{R} f_{1,m-1} + f_{0,m}$$

M_1 is the total number of sets that are not completely concordant, the values $\{k_m\}$ represent the permutations of possible values for the number of matched case-control sets with an exposed case when there are m exposed subjects in a set,

$$k = \sum_{m=1}^{R} k_m$$

and

$$C_m = \frac{R + 1 - m}{m}$$

For mid-*P* exact confidence limits, only half the probability is included in the tail for

$$k = \sum_{m=1}^{R} f_{1,m-1}$$

These limits are the solution to the equations

$$\alpha/2 = \frac{1}{2} \sum_{k=a}^{M_1} \prod_{m=1}^{R} \binom{f_{1,m-1} + f_{0,m}}{k_m} \left(\frac{OR}{OR + C_m}\right)^{k_m} \left(\frac{C_m}{OR + C_m}\right)^{f_{1,m-1} + f_{0,m} - k_m} + \sum_{k=a+1}^{M_1} \prod_{m=1}^{R} \binom{f_{1,m-1} + f_{0,m}}{k_m} \left(\frac{OR}{OR + C_m}\right)^{k_m} \left(\frac{C_m}{OR + C_m}\right)^{f_{1,m-1} + f_{0,m} - k_m} \quad [13\text{-}34]$$

and

$$1 - \alpha/2 = \frac{1}{2} \sum_{k=a}^{M_1} \prod_{m=1}^{R} \binom{f_{1,m-1} + f_{0,m}}{k_m} \left(\frac{\overline{OR}}{\overline{OR} + C_m}\right)^{k_m} \left(\frac{C_m}{\overline{OR} + C_m}\right)^{f_{1,m-1} + f_{0,m} - k_m} + \sum_{k=a+1}^{M_1} \prod_{m=1}^{R} \binom{f_{1,m-1} + f_{0,m}}{k_m} \left(\frac{\overline{OR}}{\overline{OR} + C_m}\right)^{k_m} \left(\frac{C_m}{\overline{OR} + C_m}\right)^{f_{1,m-1} + f_{0,m} - k_m} \quad [13\text{-}35]$$

In considering exact hypothesis testing, we saw that for the data in Example 13-1 (Table 13-7) there were three outcomes, including the observed data, that give $k = 11$, and only one more extreme outcome, for which $k = 12$. Therefore, $a = 11$ and $M_1 = 12$. The four terms in the summation of the upper-tail probability are

$$P_1 = \binom{4}{3}\left(\frac{OR}{OR+4}\right)^3\left(\frac{4}{OR+4}\right)^1\binom{5}{5}\left(\frac{OR}{OR+3/2}\right)^5\binom{3}{3}\left(\frac{OR}{OR+2/3}\right)^3$$

(the probability of the observed data),

$$P_2 = \binom{4}{4}\left(\frac{OR}{OR+4}\right)^4\binom{5}{4}\left(\frac{OR}{OR+3/2}\right)^4\left(\frac{3/2}{OR+3/2}\right)^1\binom{3}{3}\left(\frac{OR}{OR+2/3}\right)^3$$

$$P_3 = \binom{4}{4}\left(\frac{OR}{OR+4}\right)^4\binom{5}{5}\left(\frac{OR}{OR+3/2}\right)^5\binom{3}{2}\left(\frac{OR}{OR+2/3}\right)^2\left(\frac{2/3}{OR+2/3}\right)^1$$

and

$$P_4 = \binom{4}{4}\left(\frac{OR}{OR+4}\right)^4\binom{5}{5}\left(\frac{OR}{OR+3/2}\right)^5\binom{3}{3}\left(\frac{OR}{OR+2/3}\right)^3$$

(the probability of the most extreme outcome).

For the exact Fisher lower confidence limit, trial values of the odds ratio should be substituted in these expressions until the sum of all four terms equals $\alpha/2$. The exact Fisher upper limit is obtained by finding the value of the odds ratio for which P_4 alone is $1 - \alpha/2$. (The remaining terms, P_1, P_2, and P_3 are included in the other tail for the Fisher upper limit.) If α is chosen to be 10 percent, the exact Fisher 90 percent confidence limits are found to be 3.78 for the lower limit and 496 for the upper limit.

To get the exact mid-*P* confidence limits, the same procedure is used except that only half the probabilities calculated for P_1, P_2, and P_3 are included in the summation: the lower limit is the value of the odds ratio for which $\frac{1}{2}(P_1 + P_2 + P_3) + P_4 = \alpha/2$, and the upper limit is the value for which $\frac{1}{2}(P_1 + P_2 + P_3) + P_4 = 1 - \alpha/2$. By trial and error, solutions for the 90 percent exact mid-*P* limits can thus be calculated to be 4.59 and 252. The large discrepancy between the two upper exact limits, the Fisher and the mid-*P*, results from the fact that there is only one possible outcome more extreme than that actually observed; the difference between an upper boundary of 252 and 496 for the confidence interval will hardly affect the epidemiologic interpretation.

Approximate confidence limits for the maximum likelihood estimate of the odds ratio with R matched controls can be obtained from formulas 13-28 and 13-29, extended to R controls. The variance formula [Miettinen, 1970] that represents the extension of expression 13-27 to R controls is

$$\text{Var}[\ln(\widehat{OR})] \doteq \frac{1}{\widehat{OR}\sum_{m=1}^{R}\frac{(f_{1,m-1} + f_{0,m})C_m}{(\widehat{OR} + C_m)^2}} \qquad [13\text{-}36]$$

where $\widehat{OR}$ is the conditional maximum likelihood estimate of the odds ratio, C_m is $(R + 1 - m)/m$, and the other notation follows Table 13-6. Approximate confidence limits are calculated as

$$\underline{OR} = \exp\left[\ln(\widehat{OR}) - Z \cdot SE[\ln(\widehat{OR})]\right] \qquad [13\text{-}37]$$

and

$$\overline{OR} = \exp\left[\ln(\widehat{OR}) + Z \cdot SE[\ln(\widehat{OR})]\right] \qquad [13\text{-}38]$$

where the standard error is the square root of the variance expression 13-36.

For the data in Example 13-1, the maximum likelihood estimate of the odds ratio is 22.6, and the variance of the logarithm of the odds ratio is

$$\text{Var}[\ln(\widehat{OR})] \doteq \frac{1}{22.6\left[\frac{4\cdot4}{(22.6+4)^2} + \frac{5\cdot(3/2)}{(22.6+3/2)^2} + \frac{3\cdot(2/3)}{(22.6+2/3)^2}\right]} = \frac{1}{0.8864}$$

$$= 1.1282$$

which gives an approximate 90 percent confidence interval, from equations 13-37 and 13-38, of

$$\underline{OR} = \exp[\ln(22.6) - 1.645\sqrt{1.1282}] = 3.94$$

and

$$\overline{OR} = \exp[\ln(22.6) + 1.645\sqrt{1.1282}] = 130$$

The lower limit is a reasonably good approximation to the mid-*P* exact lower limit, but the upper limit is notably off. Nevertheless, there is no material difference in interpretability for upper limits of 130 and 252, so there is no practical problem despite the apparent inaccuracy. Discrepancies of this magnitude from different approaches to interval estimation are not unusual when the upper limit of the odds ratio is extremely high; it is important to remember that for these data the point estimate of the odds ratio is 22.6, itself an extremely high value.

For the Mantel-Haenszel estimator of the odds ratio, the variance formula is given by expression 12-58 [Robins et al., 1986]. The components

of formula 12-58, when applied to matched data with a fixed R-to-1 matching ratio, are

$$\sum_i P_i R_i = \sum_{m=1}^{R} f_{1,m-1} \frac{(R + 2 - m)(R + 1 - m)}{(R + 1)^2} \qquad [13\text{-}39]$$

$$\sum_i Q_i R_i = \sum_{m=1}^{R} f_{1,m-1} \frac{(m - 1)(R + 1 - m)}{(R + 1)^2} \qquad [13\text{-}40]$$

$$\sum_i P_i S_i = \sum_{m=1}^{R} f_{0,m} \frac{m(R - m)}{(R + 1)^2} \qquad [13\text{-}41]$$

$$\sum_i Q_i S_i = \sum_{m=1}^{R} f_{0,m} \frac{m(m + 1)}{(R + 1)^2} \qquad [13\text{-}42]$$

$$\sum_i R_i = \sum_{m=1}^{R} f_{1,m-1} \frac{R + 1 - m}{R + 1} \qquad [13\text{-}43]$$

$$\sum_i S_i = \sum_{m=1}^{R} f_{0,m} \frac{m}{(R + 1)} \qquad [13\text{-}44]$$

Applying this formula to the data of Example 13-1, for which $\widehat{OR}_{MH} = 33$, the variance is calculated to be 1.5179, for a 90 percent confidence interval of

$$\exp[\ln(33) \pm 1.645 \sqrt{1.5179}] = 4.35, 250$$

The variance estimate of 1.5179 is larger than the corresponding variance estimate for the maximum likelihood estimator, which might be expected in view of the extreme departure from the null state. It is only in the vicinity of the null condition that the Mantel-Haenszel estimator is as efficient as the conditional maximum likelihood estimator.

Another approach to approximate interval estimation of the odds ratio for R controls matched to each case is the test-based procedure. As usual, these limits are

$$\widehat{OR}^{(1 \pm Z/\chi)}$$

where the χ is the result from expression 13-15. In principle, the test-based limits could be used with either the maximum likelihood estimate or the Mantel-Haenszel estimate as the anchor point. For the data in Example 13-1, the χ is 4.0, which gives a 90 percent confidence interval of 6.3 to 81 when the maximum likelihood estimate of 22.6 is used as the anchor point, and 7.8 to 139 when the Mantel-Haenszel estimate of 33 is used as the anchor point. In either case the test-based limits are evidently much too narrow and would not serve as an adequate approximation to the exact confidence limits. The test-based limits are usually adequate in the vicinity of the null value of the odds ratio, but for these data, which depart strongly from the null condition, the test-based limits are a poor approximation.

INTERVAL ESTIMATION FOR A VARYING NUMBER OF CONTROLS MATCHED TO EACH CASE

With a varying number of matched controls, the probability expression for the data as a function of the odds ratio is an extension of formula 13-31, taking the product of the probabilities over each value of R:

$$\Pr(\text{data}) = \prod_R \prod_{m=1}^{R} \binom{f_{1,m-1} + f_{0,m}}{f_{1,m-1}} \left(\frac{OR}{OR + C_m}\right)^{f_{1,m-1}} \left(\frac{C_m}{OR + C_m}\right)^{f_{0,m}} \qquad [13\text{-}45]$$

where the notation is that used for equation 13-31. The tail probabilities for the calculation of exact confidence limits are calculated as they are for a fixed matching ratio (formulas 13-32 through 13-35) with expression 13-45 representing the probability for each realization of the data in the tail summation.

The data in Example 13-2, for which there are 3522 terms in the tail summation, would not ordinarily warrant an exact calculation of confidence limits because the large numbers ensure that most approximate formulas for the determination of confidence intervals would be satisfactory. The exact confidence limits must be determined iteratively, so that the tail summation involving 3522 terms must be calculated repeatedly until the solution is reached. This tedious task is not difficult, however, using a computer. The 90 percent exact Fisher confidence limits for the data in Example 13-2 are 1.28 and 3.10; the 90 percent exact mid-*P* limits are 1.32 and 3.01.

Approximate confidence limits for the conditional maximum likelihood estimate of the odds ratio with a varying number of matched controls can be based on formulas 13-37 and 13-38 after extending the variance formula (13-36) to accommodate more than one value for R, by extending the summation in the denominator of formula 13-36 to the various values for R:

$$\text{Var}[\ln(\widehat{OR})] \doteq \frac{1}{\widehat{OR} \sum_R \sum_{m-1}^{R} \frac{(f_{1,m-1} + f_{0,m})C_m}{(\widehat{OR} + C_m)^2}} \qquad [13\text{-}46]$$

where the notation follows that in formula 13-36.

For the data in Example 13-2, the maximum likelihood estimate of the odds ratio is, from equation 13-8, 1.9835. Substituting this value for $\hat{OR}$ in equation 13-46 along with the observed frequencies gives, for the variance of the logarithm of the odds ratio and approximate 90 percent confidence interval,

$$\hat{r}[\ln(\hat{OR})] \doteq \frac{1}{1.9835\left[\frac{11}{(1.9835+1)^2}+\frac{38(2)}{(1.9835+2)^2}+\frac{26\,(\frac{1}{2})}{(1.9835+\frac{1}{2})^2}\right]} = 0.0620$$

and

$$\underline{OR} = \exp[\ln(1.9835) - 1.645\sqrt{0.062}] = 1.32$$

$$\overline{OR} = \exp[\ln(1.9835) + 1.645\sqrt{0.062}] = 2.99$$

As one would expect with these moderately large numbers, these approximate confidence limits are extremely close to the mid-*P* exact limits of 1.32 and 3.01.

The Mantel-Haenszel estimator for the data of Example 13-2 is 2.062. The variance of the Mantel-Haenszel estimator for a varying ratio R of controls to cases can be obtained from formula 12-58 by extending the components of 12-58 given in formulas 13-39 through 13-44 for all values of R. Thus, each of the six summations should be summed for all values of R. For the data of Example 13-2, the variance of the Mantel-Haenszel estimator can be calculated in this way as 0.0659, which is slightly greater than the maximum likelihood variance estimator of 0.0620. The approximate 90 percent confidence limits for the Mantel-Haenszel estimator for the data of Example 13-2 are

$$\underline{OR}_{MH} = \exp[\ln(2.062) - 1.645\sqrt{0.0659}] = 1.35$$

$$\overline{OR}_{MH} = \exp[\ln(2.062) + 1.645\sqrt{0.0659}] = 3.14$$

Test-based approximate confidence limits can also be applied when the matching ratio varies, subject to the usual caution that their accuracy suffers according to how much the data depart from the null condition. Whereas test-based limits were a poor approximation for the data of Example 13-1, which indicated a strong effect, one might reasonably expect a better performance for the data of Example 13-2, which depart only modestly from the null state. For these data, the χ from formula 13-18 is 2.79. Using the maximum likelihood point estimate of 1.98, the test-based 90 percent confidence limits are 1.32 and 2.96, which are nearly identical to the interval obtained using the variance expression for the logarithm of the maximum likelihood estimator and nearly identical to the mid-*P* exact limits. Using the Mantel-Haenszel point estimate of 2.06, the test-based 90 percent confidence limits are 1.35 and 3.16, which are close to the results using the variance formula of Robins et al. [1986].

MATCHED FOLLOW-UP STUDIES

Matching can achieve in follow-up studies what it cannot achieve in case-control studies: It can prevent confounding. The crude risk comparisons from a matched follow-up study are unbiased with respect to the matching factors because of the absence of an association between exposure and the matching factors among the study subjects at the start of follow-up.

Despite this efficacy, matched follow-up studies are rare. The main reason is the great expense of matching large cohorts; follow-up studies ordinarily have many more subjects than case-control studies, and matching is usually a time-consuming process. Walker [1982] has suggested a method to improve this poor cost efficiency in matched follow-up studies by limiting data collection on unmatched confounders to those sets in which an event occurs. Another reason that matched follow-up studies are rare is that matching can reasonably be accomplished only for subjects themselves, whereas in any long-term follow-up study the optimal measure to use for follow-up experience is person-time. If matching were employed in a long-term follow-up study at the time of subject selection, the identical distributions of the compared series for the matched factors could change as the follow-up experience of the compared groups began to differ.

For matched follow-up studies in which the period of follow-up is short enough to warrant the use of cumulative incidence data rather than incidence rate data, a crude analysis of the data will give results that are unconfounded by the matching factors (although the crude analysis will yield a variance estimate that is too large [Greenland and Robins, 1985a]). In addition to preventing confounding, matching also contributes to study efficiency by reducing the variation of the effect estimate; the reduced variation stems from the correlation in the disease outcome for the matched subjects introduced by the matching.

Consider a matched follow-up study with T matched pairs of exposed and unexposed subjects. Suppose that the frequency distribution of matched pairs according to the outcome in exposed and unexposed subjects is f_{11} for pairs in which both the exposed and unexposed subjects develop the disease, f_{10} for pairs in which only the exposed subject develops the disease, f_{01} for pairs in which only the unexposed subject develops the disease, and f_{00} for pairs in which neither subject develops the disease. The risk difference can be estimated by

$$\hat{RD} = \hat{R}_1 - \hat{R}_0 = \frac{f_{11}+f_{10}}{T} - \frac{f_{11}+f_{01}}{T} = \frac{f_{10}-f_{01}}{T} \qquad [13\text{-}47]$$

nd the risk ratio can be estimated as

$$\widehat{RR} = \frac{(f_{11} + f_{10})/T}{(f_{11} + f_{01})/T} = \frac{f_{11} + f_{10}}{f_{11} + f_{01}} \qquad [13\text{-}48]$$

Statistical hypothesis testing for these data is identical to the procedures ısed for case-control data; both the exact and approximate methods apply :qually well for follow-up data in which all of the observations are fre-quencies. Exact confidence limits for the above measures are difficult to obtain, but excellent approximate methods exist that take into account the reduced variation introduced by the matching.

The most direct approach involves variance formulas corresponding to the estimators given in formulas 13-47 and 13-48. For the rate difference estimate, the variance is

$$\text{Var}(\widehat{RD}) \doteq \frac{T(f_{10} + f_{01}) - (f_{10} - f_{01})^2}{T^3} \qquad [13\text{-}49]$$

The variance estimate for the logarithmically transformed rate ratio measure is

$$\text{Var}[\ln(\widehat{RR})] \doteq \frac{f_{10} + f_{01}}{(f_{11} + f_{10})(f_{11} + f_{01})} \qquad [13\text{-}50]$$

The estimates of effect derived from formulas 13-47 and 13-48 are those obtained from the crude data, but the corresponding variances in formulas 13-49 and 13-50 are generally smaller than those obtained from a crude analysis. Another possible approach to confidence interval estimation is the use of test-based confidence limits, using the χ from formula 13-12.

Example 13-3 illustrates data from a follow-up study of 458 pregnant women who had previously used oral contraceptives; the comparison

Example 13-3. Distribution of matched pairs of pregnant women exposed and unexposed to oral contraceptives according to selected abnormalities in the offspring [Robinson, 1971]

	Unexposed mother		
	Abnormality present	Abnormality absent	Total
Exposed mother			
Abnormality present	28	85	113
Abnormality absent	61	284	345
Totals	89	369	458

Table 13-8. Crude data for example 13-3

	Oral contraceptive exposure		
	Yes	No	Total
Abnormal baby			
Yes	113	89	202
No	345	369	714
Total	458	458	916

group consists of an equal number of women who had never used oral contraceptives and who were individually matched to the exposed women for age and parity [Robinson, 1971]. The pairs are classified according to whether or not each mother delivered a baby with one of a group of abnormalities potentially related to the exposure. (The reader should note that these data, although they resemble cumulative incidence data, are actually prevalence data, since miscarriages are excluded.)

The estimate of risk difference from these data, from formula 13-47, is $(85 - 61)/458 = 0.052$. A 90 percent confidence interval may be calculated from the variance as determined by formula 13-49,

$$\text{Var}(\widehat{RD}) \doteq \frac{458(85 + 61) - (85 - 61)^2}{458^3} = 0.000690$$

giving for the confidence limits

$$0.0524 \pm 1.645\sqrt{0.000690} = 0.009, 0.096$$

It is also possible to use the test-based approach, based on the χ obtained from formula 13-12 applied to the data in Example 13-3. The χ value for this example is $(85 - 61)/\sqrt{146} = 1.986$, giving a 90 percent confidence interval for the rate difference of

$$0.0524(1 \pm 1.645/1.986) = 0.009, 0.096$$

which is essentially identical to the result obtained using formula 13-49.

It is interesting to compare these results with the confidence limits obtained from the crude data, ignoring the matching. The 2 × 2 table for the crude data is shown in Table 13-8; the cell entries for this table are the marginal totals for the pairs in Example 13-3. Using the square of formula 11-17, the variance for the risk difference is

$$\text{Var}(\widehat{RD}) \doteq \frac{(113)(345)}{458^2} + \frac{(89)(369)}{458^2} = 0.000748$$

which is somewhat larger than the variance estimate that takes the matching ratio into account. From this value a 90 percent confidence interval can be calculated as

$$0.0524 \pm 1.645\sqrt{0.000748} = 0.007, 0.097$$

The risk ratio can be estimated from the data in Example 13-3 as 113/89 = 1.27 using formula 13-48. The variance of the logarithmic transformation, taking the matching into account, is, from formula 13-50,

$$\text{Var}[\ln(\hat{RR})] \doteq \frac{85 + 61}{(113)(89)} = 0.0145$$

which gives a 90 percent confidence interval of

$$\exp[\ln(1.27) \pm 1.645\sqrt{0.0145}] = 1.04, 1.55$$

Test-based 90 percent confidence limits for the risk ratio are

$$1.27^{(1 \pm 1.645/1.986)} = 1.04, 1.55$$

which is essentially the same result. From the crude data in Table 13-8, using formula 11-18, we have

$$\text{Var}[\ln(\hat{RR})] \doteq \frac{345}{(113)(458)} + \frac{369}{(89)(458)} = 0.0157$$

which corresponds to a 90 percent confidence interval of

$$\exp[\ln(1.27) \pm 1.645\sqrt{0.0157}] = 1.03, 1.56$$

just slightly larger than the confidence intervals that take matching into account.

The analysis of matched follow-up studies with several unexposed subjects matched to each exposed subject is analogous to the analysis for paired data. The crude data provide an unbiased estimate of effect as long as the matching ratio is constant. If it varies, the methods of Chapter 12 for follow-up data should be applied, grouping the subjects into strata according to categories of the matching factor(s) to ensure control of confounding.

One of the differences between follow-up and case-control studies with respect to matching is the amount of information provided by the data about the effect of a matching factor on the disease occurrence. In a case-control study, there is no way to evaluate directly the effect of a factor that has been matched. An identical distribution in both cases and controls of any matched factor is ensured by the selection process. In a follow-up study, however, the identity of distribution is achieved for exposed and unexposed subjects before disease develops. The outcome among subjects classified at different levels of a matching factor is yet to be determined and can thus be evaluated by a straightforward comparison that is unconfounded by the exposure.

EVALUATION OF EFFECT MODIFICATION WITH MATCHED DATA

All of the estimation approaches described in this chapter involve the assumption that the effect is constant for all strata. Since the numbers within strata are usually extremely small for matched analyses, because the strata correspond to the matched sets, the usual statistical approaches to the evaluation of effect modification do not apply. It is still possible to evaluate whether the effect is constant over levels of a matching factor, however, if, for example, only a few categories of the matching factor are involved. We shall not consider this issue in detail but will discuss a simple case to demonstrate the idea.

Suppose a matched-pair case-control study were conducted with 200 pairs. Of the 200 pairs, suppose that 60 are discordant, 40 with the case exposed and 20 with the case unexposed, so that the overall estimate of the odds ratio is 2.0. The overall estimate is calculated on the assumption that the odds ratio is constant for all strata, but suppose that we want to evaluate statistically whether that is the case with respect to sex, which was one of the matching factors. To evaluate effect modification in these matched data by sex, it is necessary only to separate the discordant pairs into male and female subgroups and contrast the estimates of effect obtained from these subgroups. Of the 40 discordant pairs with an exposed case, suppose that 31 are male pairs and 9 are female, whereas 15 of the 20 discordant pairs in which the control is exposed are male pairs and 5 are female. Among males, the ratio of discordant pairs is 31/15 = 2.1, compared with 9/5 = 1.8 among females. The similarity of these ratios indicates that the data are reasonably comparable with a uniform rate ratio. A statistical test of the hypothesis that there is a uniform odds ratio for males and females amounts to a test of association of the 2 × 2 table shown in Table 13-9.

Table 13-9. Distribution of discordant pairs by exposure and gender for hypothetical matched data

	Male	Female	Totals
Case exposed	31	9	40
Control exposed	15	5	20
Totals	46	14	60

A χ test statistic for these data, from formula 11-6, gives $\chi = 0.21$, which corresponds to a *P*-value of 0.8 and is reasonably consistent with the hypothesis of a uniform effect.

More general tests of effect modification for matched data can be constructed by extending the procedure described above. Estimates of effect from several subcategories can be compared in a single test by using formula 12-60 coupled with formula 13-36 (or one of the simpler counterparts) to estimate the appropriate variances for each of the compared estimates.

UATION OF THE EFFECT OF MATCHING WITH CASE-CONTROL DATA

We have seen that the process of matching itself can introduce confounding into a case-control study whenever the matching factor is a correlate of the exposure. The confounding that is introduced becomes a substitute for whatever confounding might have been observed for the factor in the absence of matching; there would be confounding as long as the factor in question, in addition to being correlated with the exposure, is also related to disease status. If the matching factor is not related to disease status and therefore is not inherently confounding, matching for it represents overmatching because the effort of matching and the loss of efficiency in the required matched analysis do not improve the validity of the study. The matched analysis is still required even if the factor matched for would not have been a confounding factor, since matching for any correlate of exposure introduces confounding that necessitates a stratified analysis to remove it.

The penalty for matching for a factor or a set of factors that jointly are not correlated with the exposure is not as severe. If the matching factors do not introduce a correlation in the exposure histories between cases and controls, the matching has not introduced any confounding into the data, and the matched analysis need not be retained. Avoiding the matched analysis may be useful to bolster study efficiency by avoiding the loss of information from the matched case-control sets with fully concordant exposure histories, or to permit stratification by factors that have not been matched. Matching factors that are uncorrelated with the exposure in the data probably represent factors that would not have been confounding even without matching, so the matching cannot be viewed as productive, but the ability to abandon the unnecessary matched analysis in this situation mitigates the problem.

Evaluation of the relation between the matched factors and the exposure is essentially an evaluation of the confounding introduced by the matching, and it proceeds in the same way as evaluation of confounding generally. The effect estimate is calculated in two ways, by preserving the matching and ignoring it. If the effect estimate from the matched analysis differs materially from the crude estimate, that difference can usually be ascribed

to the confounding that results from the correlation between the matching factors and the exposure. The difference, if any exists, will usually be such that the crude estimate of the effect is closer to the null value than the confounded estimate, provided that the matching ratio of controls to cases is constant across sets. If no material difference exists between the crude estimate and the result of the matched analysis, then the investigator can conclude that the matching did not introduce or control any confounding, and the matching can be ignored in the analysis. It should be emphasized that the evaluation of matching, like the evaluation of confounding in general, should not be based on statistical tests but on the magnitude of the apparent bias reflected in the compared point estimates.

Consider as an example the data in Table 13-7. The maximum likelihood estimate of the relative risk from the stratified (matched) analysis is 23. The crude estimate, calculated from the crude exposure proportions of 12/18 for the cases and 16/72 for the controls, is 7.0. The discrepancy indicates that the matching factors were correlated with the exposure and therefore that the matched analysis must be retained.

If the matching ratio is constant across matched sets, the crude association between exposure and disease is usually closer to the null value than the association conditional on control of the matching factors. In unusual circumstances, however, the crude association between exposure and disease is farther from the null value than the association after stratification by the matching factors [Koepsell, 1984]. This anomaly occurs only if there is a negative correlation in exposure histories between cases and their matched controls. Ordinarily, the correlation is positive, but it may occasionally be negative either from sampling variability or from extreme effect modification. If stratification by the matching factors does lead to an effect estimate that is closer to the null value than the crude effect estimate, this result should be interpreted as a warning that the data are anomalous in some way. Koepsell has recommended an examination for effect modification in such situations; this step is generally a good idea even without the paradoxical effect of matching. It is also worthwhile verifying that no data processing or labeling errors have been overlooked.

MULTIVARIATE ANALYSIS OF MATCHED DATA

The conditional likelihood methods described in this chapter for estimating the odds ratio with individually matched data can be expressed mathematically in the form of a conditional logistic regression equation [Prentice and Breslow, 1978]. The two analytic approaches are equivalent as long as the exposure variable is dichotomous and no other factors, aside from the matching variables, are considered. The conditional logistic regression analysis does offer some advantages, however. A fundamental advantage is the ability to control conveniently for other factors that were measured but not matched for. Using the conventional stratified analysis, there is

often no way to control effectively for the matching factors and for factors not included in the matching algorithm. It is possible that, on evaluating the effect of matching and determining that it has introduced little or no confounding by the matching factors, the matched sets can be disrupted and the data stratified by factors not matched on. If, on the other hand, the matching has introduced a material correlation in the exposure histories between cases and controls, conditional logistic regression analysis (or other conditional models) allows both the removal of the confounding introduced by the matching and the control of additional unmatched confounding factors. It is also possible, although unusual, that the matching factors are confounding only conditionally on the control of unmatched confounding factors [Fisher and Patil, 1974; Miettinen, 1974], a situation that conditional logistic analysis can diagnose and deal with effectively. The latter method also allows the evaluation of exposure at several levels simultaneously, a process that is otherwise especially difficult with matched data (see Chap. 16). The drawback of the multivariate approach is the requirement for a computer and the necessary software for the analysis; the conventional stratified approach requires only a pencil and paper and perhaps a pocket calculator.

The construction of the multivariate model for logistic regression analysis is discussed in the next chapter, along with the theory behind the approach.

REFERENCES

Breslow, N. Odds ratio estimators when the data are sparse. *Biometrika* 1981;68:73–84.

Fisher, L., and Patil, K. Matching and unrelatedness. *Am. J. Epidemiol.* 1974;100:347–349.

Gordis, L. Should dead cases be matched to dead controls? *Am. J. Epidemiol.* 1982;115:1–5.

Greenland, S. The effect of misclassification in matched pair case-control studies. *Am. J. Epidemiol.* 1982;116:402–406.

Greenland, S., and Robins, J. M. Estimation of a common effect parameter from sparse follow-up data. *Biometrics* 1985a;41:55–68.

Greenland, S., and Robins, J. M. Confounding and misclassification. *Am. J. Epidemiol.* 1985b;122:495–506.

Jick, H., Miettinen, O. S., Neff, R. K., et al. Coffee and myocardial infarction. *N. Engl. J. Med.* 1973;289:63–77.

Koepsell, T. Unmatched analysis of matched data: Is the bias always conservative? *Am. J. Epidemiol.* 1984;120:463–464.

Lubin, J. H. An empirical evaluation of the use of conditional and unconditional likelihoods for case-control data. *Biometrika* 1981;68:567–571.

McNemar, Q. Note on the sampling of the difference between corrected proportions or percentages. *Psychometrika* 1947;12:153–157.

Miettinen, O. S. Individual matching with multiple controls in the case of all-or-none responses. *Biometrics* 1969;25:339–355.

Miettinen, O. S. Estimation of relative risk from individually matched series. *Biometrics* 1970;26:75–86.

Miettinen, O. S. Confounding and effect modification. *Am. J. Epidemiol.* 1974;100:350–353.

Miettinen, O. S. Stratification by a multivariate confounder score. *Am. J. Epidemiol.* 1976;104:609–620.

Poole, C. Exposure opportunity in case-control studies. *Am. J. Epidemiol.* 1986;123:352–358.

Prentice, R. L., and Breslow, N. E. Retrospective studies and failure time models. *Biometrika* 1978;65:153–158.

Robins, J. M., Breslow, N., and Greenland, S. Estimators of the Mantel-Haenszel variance consistent in both sparse data and large strata limiting models. *Biometrics* 1986;42:311–323.

Robinson, S. C. Pregnancy outcome following oral contraceptives. *Am. J. Obstet. Gynecol.* 1971;109:354–358.

Walker, A. M. Efficient assessment of confounder effects in matched follow-up studies. *Appl. Stat.* 1982;31:293–297.

Wilson, E. B. Probable inference. The law of succession and statistical inference. *J. Am. Stat. Assoc.* 1927;22:209–211.