MATCHING IN EPIDEMIOLOGIC STUDIES: VALIDITY AND EFFICIENCY CONSIDERATIONS

by

Donald K. Lewis

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series #1354

August 1981

MATCHING IN EPIDEMIOLOGIC STUDIES: VALIDITY AND EFFICIENCY CONSIDERATIONS

by

Donald K. Lewis

A Dissertation submitted to the faculty of The University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill

1981

Approved by:

Laurence L Kupper
Advisor

David G. Kleinbaum
Reader

Roger Grimson
Reader

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I am indebted to the chairman of my dissertation committee, Dr. Lawrence L. Kupper, for suggesting this topic and providing encouragement and expert direction throughout the course of the work. Thanks are also due to the other members of my committee: Drs. Roger Grimson, Harvey Checkoway, David Kleinbaum, and P.K. Sen. Special thanks go to Dr. Kleinbaum for his comments on Chapter 4 and to Dr. Sen for his asstiance with Chapter 3.

I am also grateful for the guidance of my program advisor, Dr. Michael J. Symons, whose constant encouragement during the early years of my graduate training played an integral role in my decision to pursue the doctoral degree.

For the consistent support of my wife, Cindy, I am especially grateful. Her commitment to me never wavered even as she assumed the burden of writing her own dissertation and the added responsibilities of being a mother to our new-born daughter, Ashley.

Thanks are due to Jackie O'Neal for her expert typing of the manuscript.

Support for my graduate training was made possible by NIOSH grant #5-T15-OH07101-04 for Occupational Health Training in Biostatistics.

LIST OF TABLES

TABLE

TABLE

TABLE

LIST OF FIGURES

DONALD K. LEWIS. Matching In Epidemiologic Studies: Validity and Efficiency Considerations. (Under the direction of LAWRENCE L. KUPPER.)

ABSTRACT

Matching is a popular method for choosing referent or control subjects in epidemiologic studies. Although matching for many years was considered primarily as a method to control confounding, recent work has suggested that the main advantage of matching over random sampling obtains from frequent increases in efficiency which can result from analyses of matched data. There nevertheless remains a controversy regarding the mertis of matching and the appropriate circumstances in which matching should be implemented for subject selection.

The primary intent of this work is to extend the methodology of Kupper, et al. (1980) to the case of several potential confounding variables. In this context conditions for no confounding are derived, and issue relating to the presence and control of confounding are discussed.

Evaluation of the relative efficiency of (frequency-) matching and random sampling is conducted by constructing ratios of "expected" Mantel-Haenszel X^2 statistics involving parameters from probabilistic population models. For local alternatives these ratios are shown to be asymptotically equivalent to the Pitman (relative) efficiency (Gibbons, 1971). Attention is restricted to dichotomous disease, exposure, and extraneous variables. Particular issues which are addressed include: (a) the relationship between the relative efficiency and the nature of the underlying confounding; (b) whether there is a loss in efficiency

from matching on non-confounders; (c) the effect of loss of sample size in the process of matching on the relative efficiency; (d) the dynamics of matching on two potential confounders which themselves may be correlated; and (c) the relative merits of pair- and frequency-matching under the circumstances of categorical matching variables.

CHAPTER 1

INTRODUCTION AND REVIEW OF THE LITERATURE

1.1 Introduction

The past thirty years has been a period of increasing interest in and utilization of observational studies, both in the health and social sciences. In the field of epidemiology this method of inquiry has indeed emerged as the modus operandi by which the relationship between a particular health effect and its hypothesized determinants is examined.

The observational study is usually characterized by the violation of one of the necessary conditions for a controlled experiment outlined by Wold (1956), in particular, the presence of uncontrolled variation due to the failure to randomize. Cochran (1965) extended this distinction by adding that the comparison groups in the observational study are subject to different "treatments" which are preassigned in a non-random manner. In the absence of randomization the greatest concern to the researcher employing the observational study method is the extrication of the effects of the disease determinant from those of disturbing extraneous variables, known as "confounders." As McKinlay (1975a) has recognized, the outstanding technique among those available to remove this type of bias has been matching.

It should be of no surprise that matching has gained widespread use in observational studies as the principal tool for bias reduction. Matching originated as a method to reduce variation in experimental studies, where homogeneity of comparison groups were ensured by random subject allocation. As observational studies and controlled experiments are similar with respect to principles of investigation, the use of matching to ensure the comparability between subgroups in observational studies spilled over naturally from the experimental setting. While the utility of matching as a method of bias control is widely recognized, more recently it has been considered relative to other sampling methods from the point of view of efficiency, as well. Although much attention has been directed to the study of the issues surrounding this choice of subject selection in observational studies, disagreement and misunderstanding still prevail about its merits.

The purpose of this portion of the chapter is a review of the development of matching in observational studies, as it pertains to the control of bias and efficiency. The focus is restricted to the case of a single matching variable, a disease variable, and an exposure variable, all of which are dichotomous. The case of continuous matching, disease or exposure variables will not be considered in this work. For a recent treatment of continuous-variable matching see Raynor and Kupper (1977). Before proceeding to a review of the pertinent literature, terminology and notation are developed in the following sections.

1.2 Terminology

1.2.1 Matching

Matching is a method of selecting a referent (comparison) group (unexposed subjects in a follow-up study, non-diseased subjects in a case-control study) such that the referent groups is "similar" to the index group (exposed group in the follow-up study, diseased group in the case-control study) with respect to the matching variables. The nature of the similarity depends upon the type of matching utilized, pair-matching or frequency-matching. Billewicz (1965) describes the proper distinction between the two. In pair-matched samples each index subject is paired with a referent subject who exhibits identical values of the matching variables. Each pair can be regarded as a distinct observational unit defining a category or stratum of the joint distribution of the matching or confounding variables.

Frequency-matching involves manipulation of the selection of referent subjects in such a way that the distribution of the confounding variables in that group is identical to the distribution in the index group. Pairing may or may not be employed to accomplish this but if so, the identify of the pairs is not preserved. Both methods accomplish the task of equating the distributions of the confounding variables in the referent and index samples; pair-matching simply continues by forming pairs of individuals subjects within the strata defined by the confounding or matching variables. In the case of categorical matching variables McKinlay (1977) has observed the arbitrariness of this scheme. Issues raised by this phenomenon will be considered in Chapter 2.

1.2.2 Confounding

Confounding is a particular type of bias which results from the failure to isolate the effect of extraneous variables from the effect of the exposure variable. Miettinen (1975) describes confounding as the "mixture of the effect at issue and the effect - possibly spurious - of another factor associated with the one being studied." A confounder is a variable which exhibits such properties if not properly controlled for in the study design or the statistical analysis. In brief, a confounder must be related to both disease and exposure. Miettinen (1974, 1975) lists more specific properties a variable must possess before it can be labelled a confounding variable. A particular objective of this paper is the specification of the conditions for no confounding when two or more potential confounders are considered simultaneously. As matching has been regarded as a primary method for control of confounding, the effectiveness of matching in this regard will be examined vis-a-vis other effective methods of control, in particular, stratification.

1.2.3 Stratification

An alternative to matching is independent random sampling of index and referent series followed by adjustment for the effects of extraneous variables in the analysis. In their classic paper on statistical analysis of retrospective study data, Mantel and Haenszel (1959) outlined the approach, which has been subsequently referred to as stratification.

While matching is essentially a design method, usually a constraint on the selection of the referent series, stratification is strictly an

analytical method and is data-dependent. It involves the subclassification of subjects into strata based upon levels of the extraneous variables, performing the analysis within each stratum, and providing an overall summary measure of effect and/or test of significance. How the summarization is accomplished presents a dilemma. Summary measures involve weighted combinations of the stratum-specific measures, and the choice of weights is often arbitrary although rarely inconsequential. Miettinen (1972) considered various choices of weights for standardization of absolute and relative measures of effect, which is a form of stratification where the weights represent functions of the distribution (among the index group, or referent group, or both) of the confounding variables. Another choice for the weights are the inverse-variances, which are optimal for minimizing the variance of the summary statistic.

A disadvantage of standardization (stratification) is that the data may "thin out" as the number of subclassifications grows larger. A result of over-stratification may be essentially empty strata or non-overlapping strata, where there is a gross imbalance between the number of referent and index subjects. In practice one is limited by this phenomenon to relatively small numbers of variables on which to stratify. However, matching is limited in the same manner since the number of matched pairs (sample sizes) can dwindle rapidly as the number of matching variables increases.

1.3 Notation and Probabilistic Models

The underlying probabilistic framework is developed in this section. The most convenient manner in which to introduce the parameters of the models to be considered is to deal with the follow-up study and case-control study separately. For both models D will represent the disease outcome $(D,\overline{D})$, E the exposure status $(E,\overline{E})$, and F the extraneous variable (F_1,F_0), which will represent a potential confounder. For the purposes of developing the models only one confounder will be involved, although the definitions can be easily extended to two or more.

The effect measures of interest in these discussions are the two most common risk indicators: risk ratio or relative risk (RR) and risk difference (RD). In situations where neither of these parameters can be estimated, the odds ratio (OR), a surrogate measure for RR, will be considered. For a discussion and comparison of the many effect measures proposed for the 2×2 table, see a recent article by Hamilton (1979).

The general layout for either of these two study designs can be described by a 2^3 contingency table. Although the proper distributional characterizations for such models are described by independent multinomials, the parameters will be defined somewhat differently than the usual multinomial parameters.

1.3.1 Follow-up Study

In this setting subjects are selected from each of two exposure groups, E and $\overline{E}$. The following parameters will be used to describe

this population. For $i = 0,1$:

$$\alpha_i = P(D/EF_i) \ , \quad \beta_i = P(D/\overline{E}F_i)$$

$$\theta_i = P(F_i/E) \ , \quad \phi_i = P(F_i/\overline{E}) \qquad (1.3.1)$$

$$\psi = P(E) \ .$$

A typical 2×2 table from the i^{th} stratum representing the conditional cell frequencies is given below, assuming equal sample sizes of referent and index subjects.

TABLE 1.1

Expected Stratum-Specific Cell Frequencies:
Follow-up Study

F_i	E	$\overline{E}$
D	$N\alpha_i\theta_i$	$N\beta_i\phi_i$
$\overline{D}$	$N(1-\alpha_i)\theta_i$	$N(1-\beta_i)\phi_i$
	$N\theta_i$	$N\phi_i$

All of the appropriate measures of association between D and E can be defined in terms of these parameters. For the i^{th} stratum as described above, the risk ratio (RR_i) and risk difference (RD_i) are defined by

$$RR_i = \frac{P(D/EF_i)}{P(D/\overline{E}F_i)} = \frac{\alpha_i}{\beta_i}$$

$$RD_i = P(D/EF_i) - P(D/\overline{E}F_i) = \alpha_i - \beta_i \ .$$

The corresponding crude measures of effect, which are functions of the cell frequencies of the 2×2 table collapsing over the strata, are

$$cRR = \frac{P(D/E)}{P(D/\bar{E})} = \frac{\alpha_1\theta_1+\alpha_0\beta_0}{\beta_1\phi_1+\beta_0\phi_0}$$

$$cRD = P(D/E) - P(D/\bar{E}) = \alpha_1\theta_1 + \alpha_0\theta_0 - \beta_1\phi_1 - \beta_0\phi_0 .$$

Two standardized measures for the risk ratio introduced by Miettinen (1972b) are denoted the "internally standardized" risk ratio (sRR) and the "externally standardized" risk ratio (s'RR). Both of these measures can be written in terms of ratios of standardized risks, the standards being the distribution of the exposed for sRR and the distribution of the unexposed for s'RR. Hence,

$$sRR = \frac{\sum_i (N\theta_i)\alpha_i}{\sum_i (N\theta_i)\beta_i} = \frac{\alpha_1\theta_1+\alpha_0\theta_0}{\beta_1\theta_1+\beta_0\theta_0}$$

$$s'RR = \frac{\sum_i (N\phi_i)\alpha_i}{\sum_i (N\phi_i)\beta_i} = \frac{\alpha_1\phi_1+\alpha_0\phi_0}{\beta_1\phi_1+\beta_0\phi_0} .$$

Each of these measures can be considered ratios of "observed to expected" cell frequencies or conditional risks (sRR) or "expected to observed" cell frequencies or conditional risks (s'RR), as well as weighted averages of the stratum-specific risk ratios.

In a similar fashion the two corresponding standardized risk differences (sRD and s'RD) are defined by:

$$sRD = \alpha_1\theta_1 + \alpha_0\theta_0 - \beta_1\theta_1 - \beta_0\theta_1$$

$$= \theta_1(\alpha_1-\beta_1) + \theta_0(\alpha_0-\beta_0)$$

$$s'RD = \alpha_1\phi_1 + \alpha_0\phi_0 - \beta_1\phi_1 - \beta_0\phi_0$$

$$= \phi_1(\alpha_1-\beta_1) - \phi_0(\alpha_0-\beta_0)$$

1.3.2 Case-Control Study

This design is characterized by sampling from diseased and non-diseased populations, D and $\overline{D}$. The following parameters, analogous to those of the follow-up study, are defined by

$$\varepsilon_i = P(E/DF_i) \ , \quad \delta_i = P(E/\overline{D}F_i)$$

$$v_i = P(F_i/D) \ , \quad w_i = P(F_i/\overline{D})$$

$$\delta = P(D) \ .$$

"Expected" cell frequencies from the i^{th} stratum based upon sampling from these populations are given below

TABLE 1.2

Expected Stratum-Specific Cell Frequences:
Case-Control Study

F_i	E	$\overline{E}$	
D	$N\varepsilon_i v_i$	$N(1-\varepsilon_i)v_i$	Nv_i
$\overline{D}$	$N\delta_i w_i$	$N(1-\delta_i)w_i$	Nw_i

Since RR and RD are not directly estimable from case-control study data, the (exposure-) odds ratio, OR, is used to measure the disease-

exposure association. For the i^{th} stratum the odds ratio is defined by

$$OR_i = \frac{P(E/DF_i) \cdot P(\bar{E}/\bar{D}F_i)}{P(\bar{E}/DF_i) \cdot P(E/\bar{D}F_i)} = \frac{\varepsilon_i(1-\delta_i)}{\delta_i(1-\varepsilon_i)} .$$

The crude odds ratio, a function of the data from the collapsed table is

$$cOR = \frac{P(E/D) \cdot P(\bar{E}/\bar{D})}{P(\bar{E}/D) \cdot P(E/\bar{D})}$$

$$= \frac{(\varepsilon_1 v_1 + \varepsilon_0 v_0)((1-\delta_1)w_1 + (1-\delta_0)w_0)}{(\delta_1 w_1 + \delta_0 w_0)((1-\varepsilon_1)v_1 + (1-\varepsilon_0)v_0)} .$$

Miettinen (1972b) also introduced two possible methods of standardizing the odds ratio, the procedures for which closely follow the methods for standardizing the risk ratio. Both measures can be interpreted as weighted averages of the stratum-specific OR's or as ratios of observed and expected cell frequencies. Let the table below represent data from a case-control study.

TABLE 1.3

Observed Stratum-Specific Cell Frequencies

F_i	E	$\bar{E}$	
D	a_i	b_i	M_{1i}
$\bar{D}$	c_i	d_i	M_{2i}
	N_{1i}	N_{2i}	N_i

In particular, Miettinen defined as the "internally standardized" odds ratio

$$sOR = \frac{\sum_i a_i}{\sum_i b_i c_i / d_i} .$$

This measure is the ratio of the observed number of exposed among the diseased group to the expected number of the same under the conditions of "no association" ($OR_1 = OR_0 = 1$). An alternative measure was also proposed by Miettinen, the "externally standardized" OR, defined by

$$s'OR = \frac{\sum_i a_i d_i / c_i}{\sum_i b_i}$$

which can be considered as the ratio of the expected number of non-exposed among the diseased group to the observed number in that category under the same assumption of "no association."

To write these parameters in terms of the notation from Section 1.2 note that

$$sOR = \frac{\sum_i \varepsilon_i v_i}{\sum_i \frac{(1-\varepsilon_i) v_i \delta_i}{1-\delta_i}}$$

$$s'OR = \frac{\sum_i \frac{(1-\delta_i)\varepsilon_i v_i}{\delta_i}}{\sum_i (1-\varepsilon_i) v_i} .$$

Certainly, other methods of standardization of the odds ratio can be derived; however, only the above methods will be considered in this paper. The choice of standard is arbitrary although not without consequence unless there is reasonable uniformity of the stratum-specific odds ratio over the strata. One can argue that when there is substantial non-uniformity of effect, any standardization or stratification

procedure has limited appeal. Therefore, the choice of standard is not of overriding importance and will not be discussed further in this work.

1.4 Review of the Literature

1.4.1 Early Work

The first major attempt to acknowledge the problem of confounding in observational studies was made by Cochran (1953) and Greenberg (1953) in companion papers published in the Journal of the American Public Health Association. In the context of continuous response variables, Cochran considered the comparison between pair-matched samples and random samples with covariance adjustment. The comparison focused on efficiency considerations, or the expected variances of the estimators of mean difference appropriate to the two designs. Although no analytical treatment was attempted, the issue that matching might be as effective as random sampling in controlling the effects of disturbing variables was raised. Cochran did not make any formal conclusions regarding the issue. Greenberg contrasted analysis of covariance to the method of "balancing" (a method of subject selection similar to frequency-matching) in terms of loss of sensitivity of the F-test. He opted for covariance analysis over balancing in most situations. While not directly applicable to the case of dichotomous data, these papers set the stage for similar work in the categorical data framework.

The smoking-lung cancer controversy stimulated two relevant papers from the same journal of the National Cancer Institute, the more well-known by Mantel and Haenszel (1959) concerning analysis of data from

retrospective studies. They recognized the importance of avoiding spurious associations in epidemiologic studies and discussed pair-matching and stratification as appropriate procedures for controlling confounding factors. No comparison of the two methods was made although some of the relative advantages and disadvantages of each were outlined. Without any further discussion the authors noted that confounding factors are related to disease, the antecedent factor or "cause," in the retrospective study.

The important contribution of the paper was the overall stratified test of association, the well-known Mantel-Haenszel X^2 test. Briefly, the test assumes a consistent association over the strata, defined by the confounding variables, and that the marginals of each stratum (2×2 table) are fixed; hence, the hypergeometric model holds. The test statistic can be expressed as

$$T = \frac{\left[\sum_i (a_i - E(a_i))\right]^2}{\sum_i \mathrm{Var}(a_i)}, \qquad (1.4.1)$$

where a_i = the number of D's who exhibit E in the i^{th} statum

and,

$$E(a_i) = \text{hypergeometric mean}$$

$$\mathrm{Var}(a_i) = \text{hypergeometric variance}.$$

Under the hypothesis of no (consistent, uni-directional) association, T follows asymptotically the X^2 distribution with 1 degree of freedom.

Mantel and Haenszel gave thought to the logical consequence of an overall significance test — a summary measure of association (an adjusted overall odds ratio). The problem of weighting was approached, and two

criteria by which weights can be chosen were suggested. The first method would weight on the basis of the precision of the estimate for each stratum; the second method would produce weights proportional to the "importance" of the increased risk in the stratum (the larger the absolute increase in risk, the more "important" the relative risk/odds ratio). As a compromise to these two criteria the authors suggested the overall estimator

$$OR_{MH} = \frac{\sum_i a_i d_i / N_i}{\sum_i b_i c_i / N_i} \qquad (1.4.2)$$

which, they argued, produces weights favorable to both methods. An interesting property of the OR_{MH} is that it reduces to the pair-matched odds ratio estimate proposed by Kraus (1958). Though not resolved by this work, Miettinen and others resumed the development of stratified overall estimators of effect at this juncture.

1.4.2 Conditions for No Confounding

The second paper from the same issue of the NCI Journal made perhaps the first attempt at quantifying the conditions for (no) confounding. While advocating the use of a ratio measure of effect over a difference measure, Cornfield, et al. (1959) listed as one of its merits the ability to appraise the "possible non-causal nature of an agent having an apparent effect." They showed that if an agent E exhibited a sizeable cRR combining over levels of F, an extraneous variable related to E, then the relative "risk" of F with respect to E would have to exceed cRR if E indeed had no effect. This is

shown easily below. The assumptions are, using the notation of Section 1.3.1:

(a) no association between D and E conditional on $F(\alpha_1 = \beta_1, \alpha_0 = \beta_0)$;

(b) a sizable crude relative risk (cRR >>1); and

(c) a positive association between F and $E(\theta_1 > \phi_1)$.

By definition

$$cRR = \frac{\alpha_1\theta_1+\alpha_0\theta_0}{\beta_1\phi_1+\beta_0\phi_0} = \frac{\frac{\alpha_1}{\beta_1}\frac{\theta_1}{\phi_1}+\frac{\alpha_0}{\beta_1}\frac{\theta_0}{\phi_1}}{1+\beta_0\phi_0/\beta_1\phi_1} .$$

And thus,

$$cRR + cRR\,\frac{\beta_0\phi_0}{\beta_1\phi_1} = RR_{FE} + \frac{\alpha_0\theta_0}{\beta_1\phi_1} , \quad \text{where} \quad RR_{FE} = \frac{\theta_1}{\phi_1} .$$

Hence,

$$RR_{FE} = cRR + \frac{1}{\beta_1\phi_1}(\beta_0\phi_0 cRR - \alpha_0\theta_0) .$$

But

$$\beta_0\phi_0 cRR - \alpha_0\theta_0 > 0 , \text{ since } \phi_0 > \theta_0 , \; cRR > 1 , \text{ and } \alpha_0 = \beta_0 .$$

Therefore,

$$RR_{FE} > cRR .$$

By this result Cornfield, et al. essentially outlined the conditions for (no) confounding bias away from the null, a special case of general confounding. Implicit in the result, however, was the context of the follow-up study. From this work and the parallel paper by Mantel and Haenszel, possibly conflicting notions regarding the nature of the interrelationships of true confounding variables might have been supported. However, an implication of this work was that the

conditions for no confounding were at least in part a function of the study design, a point not well-understood until recently.

In a pair of related papers Bross (1966, 1967) laid down the principle that Cornfield, et al. described implicitly. He stated that "as the size of an apparent effect increases, the changes that the effect is spurious decreases." Bross's work attempted to relate an observed (but spurious) association between E and D to the properties of an extraneous variable F. In the second paper he presented a table of minimum values of the relative risk of D due to F (RR_F), which would be necessary to explain away a spurious relationship between E and D, depending on the strength of the association between E and F (ORR_{EF}). He found that when the relative risk due to E exceeds 3.0, only extreme combinations of values of RR_F and OR_{EF} can explain away the observed association. While Bross considered only confounding away from the null, it is clear from his work that he understood that the spurious nature of a D-E association is a function of both the E-F and D-F relationships.

Siegel and Greenhouse (1975) generalized Bross's work to the setting where the underlying D-E association is positive. Assuming no effect modification (uniformity of the stratum-specific risk ratio), they showed that cRR could be factored by

$$cRR = RR \cdot Bias \ , \quad \text{where} \quad RR = \frac{\alpha_1}{\beta_1} = \frac{\alpha_0}{\beta_0} \quad \text{and} \ , \quad Bias = \frac{\beta_1\theta_1+\beta_0\theta_0}{\beta_1\phi_1+\beta_0\phi_0} \qquad (1.4.3)$$

Schlesselman (1978) reported the same result and described the bias as the "spurious effect of F on the observed relative risk." Both of these results are formulations of Miettinen's (1972) factorization of the cRR. He described the factorization by

$$cRR = \hat{\rho}_s \cdot \hat{\rho}^* \qquad (1.4.4)$$

where

$\hat{\rho}_s$ = standardized morbidity (mortality) ratio with the distribution of the exposed group as the standard

$\hat{\rho}^*$ = the component attributable to confounding

$$= \frac{\text{sum of "expected" number of D's among the E's}}{\text{observed sum of D's among the E's}} .$$

Clearly, when $\hat{\rho}^* = 1$, there is no confounding bias. The measure s'RR might also be substituted for $\hat{\rho}_s$, in which case the conditions for $\hat{\rho}^* = 1$ change.

Kupper, et al. (1980) laid down explicit conditions for no confounding in follow-up studies by noting that when (1.4.3) equals 1, it follows that $(\beta_1 - \beta_0)(\theta_1 - \phi_1) = 0$. Therefore, there is confounding when extraneous variable F is associated unconditionally with E, and with D conditional on $\bar{E}$. When $\hat{\rho}_s$ = s'RR, $\hat{\rho}^* = 1$ if $(\alpha_1 - \alpha_0)(\theta_1 - \phi_1) = 0$. The same conditions for no confounding result except that the F-D association must be conditional on E. The authors emphasize that when the risk ratio is not uniform over the strata, a standardized measure may not be useful. The lack of uniformity makes any assessment of confounding somewhat academic. If, however, the assumption of uniformity is tenable, then either choice of standardized measure will result in equivalent conditions for no confounding.

The authors note that if the risk difference is the effect measure of interest, the same conditions must hold. However, uniformity of the risk ratio does not imply uniformity of the risk difference; hence, the

problem of measuring confounding must be considered only in the context of a particular effect measure of interest.

The conditions for no confounding in retrospective studies are not so clearly delineated in the literature. Schlesselman (1978) tried to apply (1.4.3) to case-control data without observing that the bias is a function of parameters which are not estimable in case-control studies. Siegel and Greenhouse (1975) also attempted to write cOR in terms of follow-up parameters. Both results are invalid.

Miettinen (1972a) extended the logic behind (1.4.4) to the case-control setting by observing that a similar factorization could describe cOR. Miettinen proposed that the components of the factorization would be

$$\hat{\rho}_s = \text{sOR} = \frac{\sum_i a_i}{\sum_i b_i c_i / d_i}$$

and,

$$\hat{\rho}^* = \frac{\sum_i b_i c_i / d_i}{\left(\sum_i b_i\right)\left(\sum_i c_i\right) / \left(\sum_i d_i\right)} ,$$

where

$$\text{cOR} = \frac{\left(\sum_i a_i\right)\left(\sum_i d_i\right)}{\left(\sum_i b_i\right)\left(\sum_i c_i\right)} . \qquad (1.4.5)$$

Kupper, et al. (1980) utilized this expression to develop explicit conditions for confounding in retrospective studies. From (1.4.5) there is no confounding if $\hat{\rho}^* = 1$. If $\hat{\rho}_s = \text{sOR}$, then (using the notation of Section 1.3.2) this condition can be reduced to:

$$[(1-\varepsilon_0)v_0(1-\delta_1)w_1 - (1-\varepsilon_1)v_1(1-\delta_0)w_0](\delta_1-\delta_0) = 0 . \qquad (1.4.6)$$

There is no confounding if:

$$\text{(a)} \quad \delta_1 = \delta_0 \quad \text{or} \quad OR_{EF/\overline{D}} = \frac{\delta_1(1-\delta_0)}{\delta_0(1-\delta_1)} = 1$$

of

$$\text{(b)} \quad \frac{(1-\varepsilon_1)v_1(1-\delta_0)w_0}{(1-\delta_1)w_1(1-\varepsilon_0)v_0} = OR_{DF/\overline{E}} = 1 \ .$$

Hence, there is no confounding if F is independent of E conditional on $\overline{D}$ or if F is independent of D conditional on $\overline{E}$. If $\hat{\rho}_s$ is chosen to be s'OR, then (1.4.6) becomes

$$[\varepsilon_1 v_1 \cdot \delta_0 w_0 - \varepsilon_0 v_0 \cdot \delta_1 w_1](\delta_1 - \delta_0) = 0 \ . \qquad (1.4.7)$$

The conditions for no confounding are then:

$$\text{(a)} \quad \delta_1 = \delta_0 \quad \text{or} \quad OR_{EF/\overline{D}} = 1$$

or,

$$\text{(b)} \quad \frac{\varepsilon_1 v_1 \delta_0 w_0}{\delta_1 w_1 \varepsilon_0 v_0} = OR_{DF/E} = 1 \ .$$

The conditions for no confounding are essentially the same except that D and F must be independent conditional on E instead of $\overline{E}$.

While the above conditions are dependent upon the choice of standardized measure, if there is uniformity ($OR_1 = OR_0 = 1$), then it follows directly that $OR_{DF/E} = OR_{DF/\overline{E}}$. Hence, the choice of standard is arbitrary if there is uniformity — the conditions for no confounding are equivalent.

1.4.3 Control of Confounding: Matching

Although matching has long been valued as one of the most effective methods for controlling bias in observational studies, only recently has the effectiveness of matching in this regard been evaluated, especially in the context of completely categorical data.

The earliest paper dealing with the control of confounding via matching for the case of categorical data was by Mathen (1963), who discussed the "reduction" in bias accomplished by matching. The context of the paper is a follow-up design with the risk difference as the measure of effect.

In an empirical consideration of matching in case-control studies, Bross (1969) emphasized that the principal role of matching techniques was the minimization of troublesome artifacts or sampling biases encountered in these studies. The author suggested the need to match on factors strongly related to disease, but only for the purposes of "design efficiency."

This paper stimulated a response from Miettinen (1970), who considered in detail the implications of matching in case-control studies. He claimed that validity (the control of confounding) was served by matching only under the following circumstances: when the matching variable F is related to E unconditionally and to D conditional on E. No analytical derivation was made as a basis of this claim. One should note that he essentially identified the conditions for confounding, albeit for follow-up studies.

Siegel and Greenhouse (1975) observed the effect of matching on the control of confounding in follow-up studies. Although they

presumed pair-matching, they recognized that the general effect of matching is to restructure the referent group distribution of the extraneous variable is equivalent to that of the E's. Hence, they replaced the ϕ_i's in Table 1.1 and (1.4.3) with the corresponding θ_i's. As a result the new stratum-specific table and the bias are given by

TABLE 1.4

Expected Stratum-Specific Cell Frequencies:
Matching, Follow-up Study

F_i	E	$\overline{E}$
D	$N\alpha_i\theta_i$	$N\beta_i\theta_i$
$\overline{D}$	$N(1-\alpha_i)\theta_i$	$N(1-\beta_i)\theta_i$
	$N\theta_i$	$N\theta_i$

$$\text{bias} = \frac{\beta_1\theta_1+\beta_0\theta_0}{\beta_1\theta_1+\beta_0\theta_0} = 1$$

Matching is seen to control confounding in a follow-up study. One should note that although the matching involved pairing, the pairing was dropped in the analysis. The authors suggested a matched-pairs estimate of RR, the Kraus estimate for case-control studies (Kraus, 1958), which is invalid for estimating RR in follow-up studies. They also attempted a treatment of relative risk estimation from a case-control matched design by using the same follow-up parameters, which invalidates their results.

Whether matching will control confounding in case-control studies was examined by Kupper, et al. (1980). Since the result of matching is to arrange the distribution of F in the $\overline{D}$'s to be equivalent to

that of the D's, they noted that this amounts to replacing the w_i's with the v_i's in Table 1.2. Table 1.5 below represents the expected cell frequencies in the i^{th} stratum after frequency-matching.

TABLE 1.5

Expected Stratum-Specific Cell Frequencies:
Matching, Case-Control Study

F_i	E	$\overline{E}$	
D	$N\varepsilon_i v_i$	$N(1-\varepsilon_i)v_i$	Nv_i
$\overline{D}$	$N\delta_i v_i$	$N(1-\delta_i)v_i$	Nv_i

The authors observed that the crude odds ratio resulting from matched selection of non-diseased subjects would not generally correspond to sOR or s'OR. In particular if cORm represents the crude matched odds ratio, then

$$\text{cORm} = \frac{\left(\sum_i \varepsilon_i v_i\right)\left(\sum_i (1-\delta_i)v_i\right)}{\left(\sum_i \delta_i v_i\right)\left(\sum_i (1-\varepsilon_i)v_i\right)} \neq \text{sOR, s'OR} .$$

Consequently, matched sampling does not provide a valid estimator solely from the matching itself. In Chapter 2 the properties of cORm will be further evaluated with respect to control of confounding in case-control studies.

1.4.4 Efficiency of Matching

Because matching gained widespread use in experimental studies before emerging in observational studies, much of the focus on the

effectiveness of matching vis-a-vis other designs has been given to efficiency, or the control of random error. The attention given to efficiency has often superseded consideration of the problems of validity inherent in observational studies. However, Miettinen (1970) correctly observed that "...the validity objective so dominates the consideration of efficiency...that relative efficiency can be defined meaningfully only when both designs are valid." This is an important point that seems to have escaped many who have worked on the problem of the efficiency of matching.

Many of the early studies compared matched designs with unmatched designs, where no other form of control of bias was attempted. Stuart (1957) compared analytically the crude X^2 statistic with McNemar's X^2 statistic. Mathen (1963) considered the variance of the matched-pairs risk difference compared to the variance of the cRD. Worcester (1964) also compared the crude X^2 statistic with McNemar's X^2 statistic under the three conditions of uniform effect measures (RD, RR, and OR), noting no appreciable difference in matched over random sampling. Pike and Morrow (1970) derived essentially the same results, concluding that the pair-matched X^2 statistic will exceed the crude X^2 provided that the association between D and E in the two groups is strong; i.e., large number of concordant pairs with regard to D.

Billewicz (1965) assessed empirically the relative efficiency of pair-matched and simple random samples by using simulation techniques to compare variances of the risk difference. The results were inconclusive, and Billewicz admitted that his analysis failed to take into account the bias which remained uncontrolled in the analysis of the

randomly sampled data. Chase (1968) investigated three types of efficiency with regard to testing the equality of the risk differences. Here, the matching was assumed random. An interesting result of this work was that pair-matching compared favorably with random sampling, i.e., that efficiency could be served by pair-matching even if it were not based upon a potential confounding variable. In practice, of course, matching is rarely randomly-based, a proper criticism of this work.

Miettinen (1968) investigated the relative efficiency of matching in the context of the follow-up study, where he compared pair-matching on an extraneous variable and two independent series randomly pair-matched. Under the assumption of "no confounding," he defined relative efficiency as the ratio of inverse sample sizes which obtain equal power against a specified alternative to the null (RD = 0). For both cases the author derived the asymptotic distribution of the square root of McNemar's X^2 statistic, conditional on S, the number of discordant pairs among the matched pairs. The two sample sizes are approximated by expanding the conditional power functions in a Taylor series about the expected value of S, dropping all but the first-order terms. This work differs from earlier studies by comparing directed matching to random <u>matching</u> rather than to random <u>sampling</u>. In the final analysis, these results pertain only to the case where validity is not an issue. In addition, this work suffers from the same criticism as that of Chase's work, the illogic of random matching in observational studies.

In his paper on efficiency in follow-up studies, Miettinen (1968, 1969) asserted without proof that matching can only <u>hurt</u> design

efficiency in case-control studies. In a response to the former paper Bross (1969) attempted to provide a counter-example to this assertion and claimed that "by matching out a strong factor...it is possible to improve the chances of detecting a real — but relatively weak — relationship in a secondary factor...matching can improve design efficiency." In the counterexample Bross randomly repaired subjects who had been previously matched with respect to a "strong" variable and noted the loss in efficiency from doing so. Whether this analysis related meaningfully to the problem of efficiency is therefore questionable. Bross emphasized the need to incorporate the attainment of validity before the question of efficiency is considered, but did not pursue the point in this paper.

Little work has been devoted to the question of matching in case-control studies. Miettinen (1970) presented intuitive arguments for the belief that matching in case-control studies will result in a loss in efficiency if unnecessary for the sake of validity. Matching will influence efficiency only if the matching variable is related to exposure, and the influence will tend to be negative. Miettinen argues that matching under these circumstances will only reduce the number of discordant pairs, which tend to decrease in number with increasing association between F and E, yielding a smaller McNemar's X^2 statistic. While emphasizing that efficiency is a secondary goal, meaningful only for valid designs, he did not consider the relative efficiency of matching and another form of bias control, e.g., stratification, when validity is the (primary) concern.

McKinlay (1975) recognized that the comparison of pair-matching to independent random sampling is essentially between one method of

bias elimination to no elimination at all. She advocated a comparison between pair-matching and random sampling followed by stratification basing the efficiency considerations on the McNemar's X^2 statistic and the Mantel-Haenszel X^2 statistic. Assuming a large enough sample size her results imply that matching is more efficient only when the number of discordant pairs exceeds the number of discordant pairs in the majority of strata, a report which by appearance conflicts with that of Miettinen. McKinlay continued with Monte Carlo studies to investigate matching and stratification for a trichotomous response variable, concluding that pair-matching is not generally more efficient than stratification.

McKinlay (1977) also noticed another aspect of pair-matching which conceivably affects efficiency, the inherent non-uniqueness of pairs within a given category of the matching variable. She hypothesized that the formation of arbitrary pairs in categories of large size might affect precision and suggested as an alternative, frequency-matching followed by the usual stratified analysis for random samples.

In this regard Kupper, et al. (1980) followed these two suggestions and compared the efficiency of frequency-matching to random sampling followed by stratification for both the follow-up and case-control designs. The comparisons hinged on large sample approximations of "expected" confidence intervals for the particular effect measure of interest. The authors proposed evaluating the variance of ℓ, a weighted linear combination of the logarithms of the stratum-specific measures of effect, denoted by μ_i. The weights are chosen to minimize the variance of ℓ; therefore, the inverse variance are the weights. By this method the variance of ℓ is given by

$$\text{Var}(\ell) = \frac{\prod_{1}^{k} \sigma_i^2}{\sum_{1}^{k} \sigma_i^2}, \quad \text{where} \quad \sigma_i^2 = \text{Var}\,(\ln \mu_i) \,. \qquad (1.4.8)$$

For both the follow-up and case-control studies large-sample Taylor series approximations to σ_i^2 were utilized.

The authors considered the relative efficiency from two points of view: (a) examination of the difference between the theoretical variances (1.4.8) under the two designs, and (b) comparison of "expected" confidence intervals of the form $\exp(\ell \pm \beta_{1-\alpha/2}\sigma_\ell)$. The second approach is based upon substituting into "expected" values of ℓ and σ_ℓ^2 various numerical values of parameters of the underlying probabilistic models (e.g., Section 1.3) in order to identify conditions in which matching and/or random sampling would produce intervals which covered the null value. Results of the evaluations were compared to the asymptotically exact procedure of Cornfield (1956) and were found to be in good agreement.

The authors concluded that matching can yield more efficient analyses in certain circumstances for both studies. In follow-up studies the authors recommend matching over random sampling, as a gain in efficiency is generally expected when matching on a counfounder, and no loss in efficiency is expected when matching on a non-confounder. For case-control studies the authors give a qualified recommendation to matching, suggesting that it be used when there is confounding, small to moderate samples, intermediate values of the OR, and small to moderate exposure probabilities. They conclude that matching on a non-confounder always leads to a loss in efficiency (Miettinen, 1975)

although the loss is of no practical importance except in certain "uncommon" situations. Although a qualified recommendation, they note that the restricted conditions favoring matching are the circumstances most often encountered in the implementation of case-control studies.

Using similar underlying population models Samuels (1979, 1980) compared "expected" Mantel-Haenszel X^2 statistics rather than confidence intervals. Ratios of "expected" M-H X^2 statistics from matching and random sampling were evaluated for the case of no confounding in the former work, and the case of confounding in the latter work. Both analytical and numerical results, limited to the case-control study, were obtained. With a few minor exceptions they do not differ with those of Kupper, et al. (1980). Samuels claims that matching will lead to a gain in efficiency in follow-up studies when there is no confounding (F is unrelated to exposure) and a loss in efficiency in case-control studies when there is no confounding (F is unrelated to disease). For matching to have substantial impact in case-control studies, the exposure probabilities must vary considerably over the strata.

1.4.5 Multiple Confounding Variables

For the apparent purposes of simplicity and tractable mathematics, most of the literature already cited treats the limited case of one dichotomous confounding variable. Very little consideration has been given to the efficiency and validity of matching in studies when more than one confounding variables or a single variable which is polychotomous in involved in the analysis.

Schlesselman (1978) and Bross (1966) suggest reducing the former case into a single confounder by considering the joint "presence" and the joint "absence" of the multiple covariables as two levels of a single confounder and proceed as before. A simplistic suggestion, this procedure ignores the information provided by the interrelationships among the confounders as well as the conditional and marginal relationships of each to disease and exposure. Fisher and Patil (1974) considered the problem of choosing matching variables in a case-control study via numerical examples involving two dichotomous confounding variables. They posed the question: what is the correct definition of "unrelatedness" between two factors, F and E, when the factor F needs to be controlled for validity purposes? The definition offered, when more than one factor is under consideration, is as follows: a factor F and an exposure E are unrelated if F and E are statistically independent conditional upon the values of the other factors (except disease). Similar remarks hold for the definition of relatedness of F and D. The authors warn against deciding to choose matching variables based on the unconditional relationships alone. While no theoretical arguments were put forth, some pertinent examples were given to illustrate their reasoning.

1.5 Summary and Outline of Subsequent Work

Whether matching remains a useful enterprise for subject selection in observational studies is currently an open question. The primary condern of the researcher conducting an observational study is the attainment of validity. and matching may or may not accomplish the removal of

confounding, depending on the study type. The control of confounding can be assured at the analysis stage; therefore, it should not be the overriding issue with respect to matching. As Kupper, et al. maintain, the choice to match should be based for the most part on efficiency considerations.

As noted earlier little attention has been devoted to the choice of subject selection for epidemiologic studies in the context of an analysis which is free of confounding bias. There is a need for a fuller understanding of the efficiency of matching with respect to random sampling in this context. A major portion of this research is directed to evaluating the relative efficiency of matching and random sampling after removing confounding as an issue. Efficiency assessments are based on the asymptotic relative efficiency of "expected" Mantel-Haenszel X^2 statistics (M-H X^2 statistics which are functions of expected cell frequencies) for both matching and random sampling, which were used by Samuels (1979, 1980).

Some of the questions which are addressed in this dissertation include three which are germane to the issue of relative efficiency.

(1) Is there a definite relationship between the relative efficiency of matching and random sampling and the (underlying confounding), in particular, the direction of the confounding?

(2) When two extraneous variables are considered, how does the relative efficiency depend upon their intercorrelation, as well as relationships to disease and exposure?

(3) What is the effect of loss of sample size due to matching on the relative efficiency?

In addition, the conditions for no confounding when two dichotomous, potentially confounding variables are involved, are developed and studied. Other topics which are given attention include the proper variables, and the extent of the bias which remains after matching in case-control studies.

CHAPTER 2

PAIR-MATCHING VS. FREQUENCY-MATCHING

2.1 Introduction

From an historical perspective matching has always signified pair-matching. The distinction between pair- and frequency-matching was drawn some time ago in the literature (Cochran, 1953 and Billewicz, 1964); nevertheless, frequency-matching has not been given much attention as a method of subject selection. However, as frequency-matching is now being critically examined as a competitor to random sampling (e.g., Kupper, et al., 1980), it is also of interest to consider how the more established method of pair-matching measures up to this relative newcomer under conditions where frequency-matching will be utilized.

Pair-matching seeks to match an index subject with a referent subject on the basis of one or more confounding variables. This matched pair then becomes the observational unit in the analysis, replacing the two individual subjects. The matching process can be ignored, of course, and the individual subjects retained as the observational units in the analysis. In fact, a special case of frequency-matching (equal index and referent sample sizes) can arise by performing pair-matching and ignoring the identify of the pairs while retaining the strata.

In this chapter the analytical method of pair-matching will be contrasted to the analytical method of frequency-matching. Hence,

that which follows can be considered a comparison of retaining vs. disregarding the identity of the pairs in the analysis of pair-matched data. The matching variables will be considered categorical, the levels of which may be real or representative. Representative levels may result from the categorization of continuous variables, although there are methods available to match on continuous variables without having to categorize. Matching on the basis of the nearest-neighbor or minimum-distance criterion is an alternative method in which to pair subjects on continuous variables. However, it can be argued that in practice, for reasons of convenience and simplicity, the common treatment of continuous variables is to stratify them in the analysis.

2.2 Analysis of Pair-Matched Data

Consider a potential confounding variable F whose L levels are denoted by F_k, k = 1,...,L. These levels can represent the joint levels of multiple variables as well as those of a single variable. For the case of dichotomous disease and exposure variables the random variable which represents a pair of index and referent subjects is a two-dimensional multinomial random variable. The four possible outcomes of this random variable correspond to the joint presence or absence of the outcome characteristics (disease in a follow-up study and exposure in a case-control study) associated with the matched subjects. Assuming that there is a total of n pairs, let $(n_{11}, n_{10}, n_{01}, n_{00})$ be the multinomial random variable with probability vector $(\rho_{11}, \rho_{10}, \rho_{01}, \rho_{00})$, where

n_{ij} = # of matched pairs whose index subject exhibits outcome characteristic i and whose referent subject exhibits outcome characteristic j

ρ_{ij} = P (outcome characteristic i conditional on index status, outcome characteristic j conditional on referent status).

For either study-type if the matching is performed at each of the L levels of F, then the above definitions accommodate this situation with the addition of the subscript k.

2.2.1 Follow-up Study

In a follow-up study the data arising from matching within a given stratum are arranged in the 2×2 table below

TABLE 2.1

Stratum-Specific Pair-Matching Data: Follow-up Study

F_k		$\bar{E}$		
		D	$\bar{D}$	
E	D	n_{11k}	n_{10k}	$n_{1\cdot k}$
	$\bar{D}$	n_{01k}	n_{00k}	$n_{0\cdot k}$
		$n_{\cdot 1k}$	$n_{\cdot 0k}$	n_k

where

n_k = # of pairs (exposed) in k^{th} stratum

$= n\theta_k$.

Pooling over the L strata yields the table of data below.

TABLE 2.2

Pooled Pair-Matching Data: Follow-up Study

E \ $\bar{E}$	D	$\bar{D}$	
D	n_{11}	n_{10}	$n_{1\cdot}$
$\bar{D}$	n_{01}	n_{00}	$n_{0\cdot}$
	$n_{\cdot 1}$	$n_{\cdot 0}$	n

The multinomial probabilities, ρ_{ijk}, can be expressed in terms of the parameters of Section 1.3.1 as follows

$$\rho_{11k} = \alpha_k\beta_k\theta_k \qquad \rho_{10k} = \alpha_k(1-\beta_k)\theta_k$$

$$\rho_{01k} = (1-\alpha_k)\beta_k\theta_k \qquad \rho_{00k} = (1-\alpha_k(1-\beta_k)\theta_k \ .$$

The stratum-specific risk ration, RR_k is expressed as the following function of these parameters

$$RR_k = \frac{\alpha_k}{\beta_k} = \frac{\alpha_k\beta_k\theta_k+\alpha_k(1-\beta_k)\theta_k}{\alpha_k\beta_k\theta_k+\beta_k(1-\alpha_k)\theta_k} = \frac{\rho_{11k}+\rho_{10k}}{\rho_{11k}+\rho_{01k}} \ . \qquad (2.2.1)$$

It is easily shown that the maximum likelihood estimator (MLE) of RR_k under matching is given by

$$RR_k = \frac{n_{11k}+n_{10k}}{n_{11k}+n_{01k}} = \frac{n_{1\cdot k}}{n_{\cdot 1k}} \ . \qquad (2.2.2)$$

Consider the parameter

$$\eta = \frac{\rho_{11}+\rho_{10}}{\rho_{11+}\rho_{01}}$$

$$= \frac{\sum_k(\rho_{11k}+\rho_{10k})}{\sum_k(\rho_{11k}+\rho_{01k})}$$

$$= \frac{\sum_k \alpha_k \theta_k}{\sum_k \beta_k \theta_k} = sRR \ . \qquad (2.2.3)$$

Corresponding to the stratum-specific estimator, the MLE of η under matching is

$$\hat{\eta} = \frac{n_{11}+n_{10}}{n_{11}+n_{01}} \ , \qquad (2.2.4)$$

the usual pain-matched estimate of RR. Hence, pair-matching always leads to a valid estimate of RR in a follow-up study.

2.2.2 Case-Control Study

The multinomial probabilities associated with the case-control design can be expressed in terms of the parameters of Section 1.3.2, as follows:

$$\rho_{11k} = \varepsilon_k \delta_k v_k \qquad \rho_{10k} = \varepsilon_k (1-\delta_k) v_k$$

$$\rho_{01k} = (1-\varepsilon_k)\delta_k v_k \qquad \rho_{00k} = (1-\varepsilon_k)(1-\delta_k) v_k \ .$$

The effect measure of interest, the odds ratio, OR, is defined for a particular stratum of F by

$$OR_k = \frac{\rho_{10k}}{\rho_{01k}} = \frac{\varepsilon_k (1-\delta_k)}{\delta_k (1-\varepsilon_k)} \ . \qquad (2.2.5)$$

Data arising from a case-control study and pooled over the strata of F are displayed in the table below:

TABLE 2.3

Pooled Pair-Matching Data: Case-Control Study

		$\bar{D}$		
		E	$\bar{E}$	
D	E	n_{11}	n_{10}	$n_{1\cdot}$
	$\bar{E}$	n_{01}	n_{00}	$n_{0\cdot}$
		$n_{\cdot 1}$	$n_{\cdot 0}$	n

The well-known estimator for the odds ratio under matching (Fleiss, 1973) is given by

$$\hat{OR} = n_{10}/n_{01} ,$$

which can also be shown to be the MLE of the parameter

$$\rho = \frac{\sum_k \varepsilon_k(1-\delta_k)v_k}{\sum_k \delta_k(1-\varepsilon_k)v_k} = \sum_k w_k \, OR_k , \qquad (2.2.6)$$

where

$$w_k = \frac{\delta_k(1-\varepsilon_k)v_k}{\sum_k \delta_k(1-\varepsilon_k)v_k} .$$

While (2.2.6) defines a weighted average of stratum-specific odds ratios, it does not correspond to either sOR or s'OR. However, it is equivalent to the Mantel-Haenszel odds ratio, referred to in Chapter 1. Note that this measure of effect provides a valid estimate of effect. At first glance it would appear that any analysis of data from Table 2.3 would be a crude analysis. However, the analysis referenced by (2.2.6) represents implicit stratified analysis, since the matching has not been ignored in the analysis. In fact, pair-matching is actually a type of stratification with the stratum size

uniformly equal to 2. This is an important aspect regarding the analysis of pair-matched data. Because of the simplicity of Table 2.3, this fact may easily be overlooked.

This section has presented a review of generally well-known aspects of the analysis of pair-matched data. The discussion now proceeds to a critical analysis of pair- and frequency-matching.

2.3 Superiority of Frequency-Matching to Pair-Matching

In the context of matching on categorical variables, one can accomplish the process of matching in two ways: via pair- or frequency-matching. It was noted that a frequency-matched sample, restricted to equal index and referent sample sizes, can be generated by forming the matched pairs and then dropping the identity of the pairs. In this section it will be demonstrated that the analysis of pair-matched data is improved when the pairs are dropped rather than retained. In particular, the pairing will be shown to be either wasteful or lead to possibly erroneous estimation. In addition, the relative efficiency of the analysis of paired data to the analysis of non-paired data will be evaluated.

2.3.1 Estimation

In the follow-up study the estimates of RR_k and RR, (2.2.2) and (2.2.3), are both functions of the marginal frequencies of Tables 2.1 and 2.2, rather than the cell frequencies alone. The marginal frequencies from those tables represent cell frequencies of Table 2.4 below:

TABLE 2.4

Marginal Frequencies of Tables 2.1 and 2.2

TABLE 2.1

	$\bar{E}$	$\bar{E}$	
D	$n_{1\cdot k}$	$n_{\cdot 1k}$	
$\bar{D}$	$n_{0\cdot k}$	$n_{\cdot 0k}$	
	n_k	n_k	$2n_k$

TABLE 2.2

	$\bar{E}$	$\bar{E}$	
D	$n_{1\cdot}$	$n_{\cdot 1}$	
$\bar{D}$	$n_{0\cdot}$	$n_{\cdot 0}$	
	n	n	$2n$

However, these cell frequencies represent the numbers of subjects (not pairs) who are matched on F_k (Table 2.1) and pooled over F (Table 2.2). Hence, these are tables of frequency-matched subjects rather than pairs. The estimates (2.2.2) and (2.2.3) then are the stratum-specific and pooled estimates of RR from frequency-matched samples. The pairing has been dropped to do the analysis — it has not been utilized.

As a result, the design has been restricted to an equal sample-size, frequency-matched sampling scheme when perhaps a more general index-to-referent sampling ratio is preferable. In a follow-up study pairing is simply a wasteful exercise and can unnecessarily restrict the design.

In the case-control study the estimates of the odds ratios are functions of the matched pairs. However, McKinlay (1977) recognized that the pairing within each stratum of F is arbitrary. The particular configuration of pairs for each stratum are not unique; in fact, there are $n_k!$ possible pairing arrangements, and therefore, $\prod_k n_k!$ possible combinations of pairs in total. Of course, not all of those pairings will yield unique arrangements of cell frequencies in Tables 2.4 and 2.5. Assuming that the cell frequencies of Table 2.4

are fixed subject to a frequency-matched sample or a pair-matched sample ignoring the pairing, there are $(1 + \min_{ij}(n_{ijk}))$ unique pairing arrangements for each stratum and $\left(1 + \min_{ij}\left(\sum_{k} n_{ijk}\right)\right)$ unique pairing arrangements for the total sample, where the n_{ijk} denote the cell frequencies of Table 2.4. As the sample size grows large, the number of potential pairing arrangements increases, as do the possible values of the estimates of effect, OR_k and OR.

This property of pair-matching seriously undermines the validity of pair-matched estimates of the odds ratio. If the pairing can be suitably manipulated at the whim of an investigator, then the estimate of effect can be so manipulated to his advantage. The pairing thus introduces an artificial source of bias to the analysis and destroys the correspondence between sampling and analysis which is required for validity in a case-control study. Note that the follow-up study does not suffer from this property because the analysis does not utilize the pairing.

The following example illustrates this thesis.

Example 1

A case-control study is performed and the subjects are selected by matching on F, a dichotomous variable. If the particular pairing which gave rise to the data is ignored, the data can be arranged in the 2×2 tables below.

TABLE 2.5

Hypothetical Pair-Matching Data: Pairing Dropped

F_1

	E	$\overline{E}$	
D	4	5	9
$\overline{D}$	2	7	9
	6	12	

$OR_1 = 28/10 = 2.80$

F_0

	E	$\overline{E}$	
D	6	9	15
$\overline{D}$	3	12	15
	9	21	

$OR_0 = 72/27 = 2.67$

The odds ratios are nearly uniform across the two strata. All of the possible pairing arrangements that could result from these data are listed below (in the form of Table 2.3).

Stratum	Pairing Arrangements			
	1	2	3	
F_1	0 4 / 2 3	1 3 / 1 4	2 2 / 0 5	
	1	2	3	4
F_0	0 6 / 3 6	1 5 / 2 7	2 4 / 1 8	3 3 / 0 9

Pooling these tables yields $\left(1 + \min_{ij}\left(\sum_{k} n_{ijk}\right)\right) = (1 + \min(10,14,5,19)) = 6$ unique configurations which can summarize the pairing. These tables and the corresponding odds ratios are listed below in Table 2.6.

TABLE 2.6

Summary of all Possible Pooled Tables

Tables (F_1,F_0)	Frequency	Pooled Table Configuration	OR	X^2_{McN}
(1,1)	1	0,10 / 5,9	$\frac{10}{5} = 2.0$	1.67
(2,1),(1,2)	2	1, 9 / 4,10	$\frac{9}{4} = 2.25$	1.92
(3,1),(1.3),(2,2)	3	2, 8 / 3,11	$\frac{8}{3} = 2.67$	2.27
(1,4),(2,3),(3,2)	3	3, 7 / 2,12	$\frac{7}{2} = 3.5$	2.78
(2,4),(3,3)	2	4, 6 / 1,13	$\frac{6}{1} = 6.0$	3.57
(3,4)	1	5, 5 / 0,14	$\frac{5}{0}$ (undef.)	5.0

There is considerable variation in the pair-matched odds ratios about the "true" odds ratio, which is between 2.67 and 2.80. Since this variation is artificially introduced only through the pairing, it is clear that the analysis of the paired data under such circumstances can lead to completely erroneous estimates of the odds ratio.

In the same way the test of significance of the pair-matched odds ratio can also be misleading. The usual test of significance for matched-pair data is McNemar's X^2_1-statistic, defined as

$$X^2_{McN} = \frac{(n_{10} - n_{01})^2}{n_{10} + n_{01}},$$

which has been shown to follow the X^2 distribution with 1 degree of freedom under the null hypothesis of no association. The test statistics for each of the six configurations are given in the fifth column of Table 2.6. The Mantel-Haenszel X^2_1-statistic based upon stratification of F in terms of the data in Table 2.5 can be shown to be 2.32. Hence, an erroneous conclusion regarding the significance of the association can result from the artificial pairing.

2.3.2 Efficiency

Beyond these arguments, the analysis of pair-matched data retaining the pairing usually results in a loss in efficiency over dropping the identity of the pairs. Let us assume that the pairing is representative of the true joint probabilities of the outcome variable (disease in a follow-up study, exposure in a case-control study) given an index and a referent subject paired on F_k, $k = 1,\ldots,L$. Then we might replace the n_{ijk} of Table 2.4 with their expected values under multinomial sampling. It is sufficient to consider only the follow-up study in this regard, as the results are completely analogous for case-control studies (due to the direct correspondence between the probabilistic model-based parameters of the two studies).

For the k^{th} stratum of F, Table 2.4 becomes

TABLE 2.7

Expected Stratum-Specific Pair-Matching Data: Follow-up Study

F_k		E	$\bar{E}$	
		D	$\bar{D}$	
E	D	$n\alpha_k\beta_k\theta_k$	$n\alpha_k(1-\beta_k)\theta_k$	$n\alpha_k\theta_k$
	$\bar{D}$	$n\beta_k(1-\alpha_k)\theta_k$	$n(1-\alpha_k)(1-\beta_k)\theta_k$	$n(1-\alpha_k)\theta_k$
		$n\beta_k\theta_k$	$n(1-\beta_k)\theta_k$	$n\theta_k$

Pooling over F the overall expected values of the cell frequencies of Table 2.4 are

TABLE 2.8

Expected Pooled Pair-Matching Data: Follow-up Study

		$\bar{E}$		
		D	$\bar{D}$	
E	D	$n\sum_k \alpha_k \beta_k \theta_k$	$n\sum_k \alpha_k (1-\beta_k)\theta_k$	$n\sum_k \alpha_k \theta_k$
	$\bar{D}$	$n\sum_k \beta_k (1-\alpha_k)\theta_k$	$n\sum_k (1-\alpha_k)(1-\beta_k)\theta_k$	$n\sum_k (1-\alpha_k)\theta_k$
		$n\sum_k \beta_k \theta_k$	$n\sum_k (1-\beta_k)\theta_k$	n

Therefore, the expected McNemar X_1^2-statistic for this table is

$$X^2_{McN} = \frac{\left(n\sum_k \alpha_k (1-\beta_k)\theta_k - n\sum_k \beta_k (1-\alpha_k)\theta_k\right)^2}{n\sum_k \alpha_k (1-\beta_k)\theta_k + n\sum_k \beta_k (1-\alpha_k)\theta_k}$$

$$= \frac{n\left(\sum_k \theta_k (\alpha_k - \beta_k)\right)^2}{\sum_k \theta_k (\alpha_k + \beta_k - 2\alpha_k \beta_k)} . \qquad (2.3.1)$$

Assuming that the pairing is dropped and the data arranged for each stratum as in Table 2.4, the expected cell frequencies can be found from the marginals of Table 2.7 above. In this case we can consider these expected cell frequencies fixed since all possible matched pairs will emanate from these marginals. For the k^{th} stratum of F, these expected frequencies are arranged below.

TABLE 2.9

Expected Stratum-Specific Cell Frequencies: Pairing Dropped, Follow-up Study

F_k		E	E	
	D	$n\alpha_k \theta_k$	$n\beta_k \theta_k$	$n\theta_k (\alpha_k + \beta_k)$
	$\bar{D}$	$n(1-\alpha_k)\theta_k$	$n(1-\beta_k)\theta_k$	$n\theta_k (2-\alpha_k-\beta_k)$
		$n\theta_k$	$n\theta_k$	$2n\theta_k$

The Mantel-Haenszel X_1^2-statistic testing the departure of these "data" from that of "no association" is expressed as

$$X_{MH}^2 = \frac{\left[\sum_k \frac{n\alpha_k\theta_k \cdot n(1-\beta_k)\theta_k - n\beta_k\theta_k \cdot n(1-\alpha_k)\theta_k}{2n\theta_k}\right]}{\sum_k \frac{n\theta_k \cdot n\theta_k \cdot n\theta_k(\alpha_k+\beta_k) \cdot n\theta_k(2-\alpha_k-\beta_k)}{(2n\theta_k)^3}} \qquad (2.3.2)$$

$$= \frac{n\left[\sum_k \theta_k(\alpha_k-\beta_k)\right]^2}{\sum_k \theta_k(\alpha_k+\beta_k)(2-\alpha_k-\beta_k)/2} \; . \qquad (2.3.3)$$

Note that (2.3.2) assumes a large sample size (replacing $2n\theta_k - 1$ with $2n\theta_k$ in the denominator). For small n we would expect X_{MH}^2 to be slightly smaller than that of (2.3.2).

Comparing (2.3.1) and (2.3.3) we note that the numerators are equal. Hence, $X_{MH}^2 > X_{McN}^2$ only if the denominator of (2.3.3) is less than that of (2.3.1); that is, only if

$$\sum_k \theta_k(\alpha_k+\beta_k)(1-(\alpha_k+\beta_k)/2) < \sum_k \theta_k(\alpha_k+\beta_k-2\alpha_k\beta_k)$$

$$-\sum_k \theta_k(\alpha_k+\beta_k)^2/2 < -2\sum_k \theta_k\alpha_k\beta_k$$

or,

$$\sum_k \theta_k(\alpha_k^2-2\alpha_k\beta_k+\beta_k^2) > 0 \iff \sum_k \theta_k(\alpha_k-\beta_k)^2 > 0 \; . \qquad (2.3.4)$$

Since $\theta_k(\alpha_k-\beta_k)^2 > 0$, $\forall$ k, (2.3.4) is always true, and therefore, the expected X_{McN}^2 is always less than the expected X_{MH}^2 statistic. The result (2.3.4) also holds for the case-control study, where θ_k, α_k, and β_k are replaced by ν_k, ε_k, and δ_k respectively.

(For small values of the measure of effect — near 1.0 — and small n, a more rigorous evaluation of the relative efficiency is presented in Appendix 1.)

The relative efficiency of the two tests can be expressed as a ratio of the two test statistics:

$$RE = \frac{\chi^2_{McN}}{\chi^2_{MH}} = \frac{\sum_k \theta_k(\alpha_k+\beta_k)(1-(\alpha_k+\beta_k)/2)}{\sum_k \theta_k(\alpha_k+\beta_k-2\alpha_k\beta_k)}$$

and can be regarded as a measure of the loss in efficiency due to the pairing. RE is defined above for the follow-up study and similarly for the case-control study with the aforementioned substitutions.

In order to gauge the magnitude of the loss from pairing, evaluations of RE were made at various values of the model-based parameters that might be expected in practical situations. Table 2.10 gives a general picture of the practical effects of pairing in the context of a follow-up and case-control study where one dichotomous matching variable is considered. In both tables the losses were averaged over confounder distributions ($\theta_1 = .1,\ldots,.9$; $v_1 = .1,\ldots,.9$) since the variation in the loss over these distributions was only slight.

By inspection of these tables it is clear that for the most part pairing does not drastically affect the efficiency of the analysis. This is especially true for the follow-up study, in which the expected loss in the most extreme case is only about 6%. In the case-control study the magnitude of the expected loss is much greater although not sizeable until the odds ratio exceeds 3.5. For situations in which only a moderate odds ratio is expected, pairing will not diminish the test statistic noticeably.

2.4 Summary

While pair-matching has historically been the usual method of matching in observational studies, if the matching variables from such a study are categorical, there is no foundation for carrying out a pair-matched analysis of the data. If matching is to be used as a method of subject selection, whether pair- or frequency-matching, the identity of the pairs should be dropped at the analysis stage when possible, and the data analyzed as if frequency-matching were performed.

In the case of the follow-up study it is shown that the measure of effect does not utilize the pairing. In case-control studies pairing can lead to a biased estimate of the odds ratio, and re-pairing of the same subjects can result in completely varying estimates of the odds ratio. For both studies pairing is less efficient than frequency-matching, although only negligibly so.

Since pairing is a form of stratification, when the matching variables are categorical, retention of the pairs becomes unnecessary over-stratification. As the number of matching categories grows larger and approaches the number of matched pairs, the problem of over-stratification diminishes as do the disadvantages of pairing summarized above.

TABLE 2.10

Expected Loss in Efficiency Due to Pairing

Follow-up Study				Case-Control Study			
RR	$\beta_0 = P(D/\bar{E}F_0)$	$OR_{DF/\bar{E}}$	Loss(%)	OR	$\delta_0 = P(E/\bar{D}F_0)$	$OR_{EF/\bar{D}}$	Loss(%)
1.50	0.001	2.0	0.01	1.50	0.01	2.0	0.08
		5.0	0.02			5.0	0.19
	0.01	2.0	0.08		0.05	2.0	0.50
		5.0	0.22			5.0	0.69
2.25	0.001	2.0	0.05	2.20	0.01	2.0	0.38
		5.0	0.10			5.0	0.82
	0.01	2.0	0.40		0.05	2.0	1.59
		5.0	1.05			5.0	2.83
3.50	0.001	2.0	0.13	3.50	0.01	2.0	1.06
		5.0	0.28			5.0	2.53
	0.01	2.0	1.16		0.05	2.0	4.36
		5.0	3.03			5.0	6.85
5.0	0.001	2.0	0.22	5.0	0.01	2.0	2.02
		5.0	0.55			5.0	4.43
	0.01	2.0	2.23		0.05	2.0	7.42
		5.0	5.85			5.0	10.51

CHAPTER 3

THE RELATIVE EFFICIENCY OF MATCHING AND RANDOM SAMPLING: ONE POTENTIAL CONFOUNDING VARIABLE

3.1 Introduction

In Chapter 1 a number of publications were cited in which the efficiency of matching was considered. In almost all of these papers matching was assumed to be pairwise, and the alternative procedure to which matching was compared involved a crude analysis of data from a random sample. For the most part, incorporating potential confounding variables into the comparisons was avoided, even though a critical element of the design and analysis of an observational study is the identification and removal of confounding bias. A notable exception is the work by McKinlay (1974), which compared pair-matching to random sampling accompanied by stratification. However, this effort stopped short of making any definitive conclusions regarding pair-matching.

In Chapter 2 pair-matching was shown to be inferior to category-matching in the context of categorical matching variables. As noted earlier, category-matching has not been studied seriously as an alternative form of matching. The work of Kupper, et al. (1980) represents a major attempt to study category-matching as a competitor to random sampling, where the control of confounding is presupposed. A key point made in that paper is that the crux of the matching/random

sampling issue pertains to the gain or loss in efficiency by matching, as opposed to the attainment of validity. Since validity can be assured by other processes, e.g., stratification, the choice of subject selection method should be based primarily on efficiency considerations. A major thrust of this work is the evaluation of the relative efficiency of matching as compared to random sampling.

3.2 Methods of Evaluation

The classical approach to the evaluation of the relative efficiency of two estimators of a parameter μ is the comparison (usually the ratio) of their asymptotic variances. Alternatives to comparing asymptotic variances include comparisons of test statistics or variances of functions of μ, if the variance of μ does not exist. Often these approaches do not yield mathematically tractable solutions, and in that case, the relative efficiency has to be studied using numerical techniques. For the most part the limited study of the relative efficiency of frequency-matching and random sampling has involved numerical approaches.

Kupper, et al. (1980) based relative efficiency comparisons on confidence intervals for difference and ratio measures of effect. Expected cell frequencies from matching and from random sampling were substituted in expressions for the variances of the risk differences and log-risk ratio in the follow-up study and the log-odds ratio in the case-control study. Treating these measures as asymptotically normally distributed, confidence intervals were constructed using Taylor-series approximations to the variances. After substituting the expected cell frequencies implied by each of the sampling designs, the intervals were compared in

terms of whether the null values of the effect measure were covered. This procedure was repeated for various values of the population parameters which might be observed in practice. The results of these comparisons were contrasted to comparisons of confidence intervals constructed using Miettinen's test-based method (1976) and Cornfield's approximate procedure (1956), which led to the same conclusions.

In much the same fashion Samuels (1979, 1980) chose to measure the efficiency of matching by considering Mantel-Haenszel X_1^2 statistics based upon the substitution of expected cell frequencies for what would be the observed values. The ratio of the "expected" test statistics from the two sampling designs was evaluated as a measure of the relative efficiency of matching to random sampling. One desirable attribute of this measure is that it can be shown to be equivalent asymptotically to the Pitman efficiency (Noether, 1950) for sequences of alternatives converging to the null hypothesis. (A proof that this index is equivalent to the Pitman efficiency is given in Appendix 3.)

In this chapter the ratio of "expected" Mantel-Haenszel X_1^2 will also be used to measure relative efficiency. This work extends that of Kupper, et al. (1980) and Samuels (1980) to a more thorough analysis of the relationship of the relative efficiency of matching to the nature and extent of (the underlying) confounding. The advantages of this approach over that of Kupper, et al. (1980) include a more refined analysis of the relative efficiency over selected regions of the population parameters. In addition, the expression of the relative efficiency is made independently of sample size.

3.3 Ratio of Expected Mantel-Haenszel X_1^2 Statistics

The expected Mantel-Haenszel X_1^2 statistics under matching and random sampling can be formed by substituting the expected cell frequencies from Tables 1.1 and 1.2, as done in Chapter 2. First, the follow-up study will be considered.

3.3.1 Follow-up Study

If the data from each stratum defined by one or more (potentially) confounding variables are arranged as in Table 1.3, then the Mantel-Haenszel X_1^2 statistic is written as:

$$X^2 = \frac{\left[\sum_i (a_i d_i - b_i c_i)/N_i\right]}{\sum_i \frac{N_{1i} N_{2i} M_{1i} M_{2i}}{N_i^2 (N_i - 1)}} \quad . \tag{3.3.1}$$

For large N_i, X^2 is adjusted slightly upward by substituting N_i^3 for $N_i^2(N_i-1)$ in the denominator. However, it will be assumed that the adjustment has only negligible effects on the relative efficiency when the level of stratification in the analysis is the same for both designs. If the expected cell frequencies of Table 1.1 are substituted in (3.3.1), then for random sampling:

$$X^2 = X_{rs}^2$$

$$= \frac{\left[\sum_i \frac{(n_1\alpha_i\theta_i)(n_0(1-\beta_i)\phi_i) - (n_1(1-\alpha_i)\theta_i)(n_0\beta_i\phi_i)}{n_1\theta_i + n_0\phi_i}\right]^2}{\sum_i \frac{(n_1\alpha_i\theta_i + n_0\beta_i\theta_i)[(n_1\theta_i + n_0\phi_i) - (n_1\alpha_i\theta_i + n_0\beta_i\phi_i)]n_1\theta_i n_0\phi_i}{(n_1\theta_i + n_0\phi_i)^3}}$$

$$= \frac{n_1 n_0 \left[\sum_i \theta_i \phi_i (\alpha_i - \beta_i)/(n_1\theta_i + n_0\phi_i)\right]^2}{\sum_i \frac{\theta_i\phi_i(n_1\alpha_i\theta_i + n_0\beta_i\phi_i)[(n_1\theta_i + n_0\phi_i) - (n_1\alpha_i\theta_i + n_0\beta_i\phi_i)]}{(n_1\theta_i + n_0\phi_i)^3}} .$$

If one can also assume that there is uniformity in the measure of effect and that a variable referent-to-index sample size ratio ($\rho = n_0/n_1$) will be allowed, then the following restrictions will hold

$$RR = \frac{\alpha_i}{\beta_i} \Rightarrow \alpha_i = RR \cdot \beta_i \ , \quad \forall_i$$

$$\rho = n_0/n_1 \Rightarrow n_0 = n_1 \cdot \rho \ .$$

The expected X^2_{rs} statistic can then be written as

$$X^2_{rs} = \frac{\rho n_1 \left[\sum \theta_i \phi_i (\alpha_i - \beta_i)/(\theta_i + \rho\phi_i)\right]^2}{\sum_i \frac{\theta_i\phi_i(\alpha_i\theta_i + \rho\beta_i\phi_i)[(\theta_i + \rho\phi_i) - (\alpha_i\theta_i + \rho\beta_i\phi_i)]}{(\theta_i + \rho\phi_i)^3}}$$

$$= \frac{\rho n_1 (RR-1)^2 \left[\sum_i \theta_i \phi_i \beta_i/(\theta_i + \rho\phi_i)\right]^2}{\sum_i \frac{\theta_i\phi_i\beta_i(RR\theta_i + \rho\phi_i)[(\theta_i + \rho\phi_i) - \beta_i(RR\theta_i + \rho\phi_i)]}{(\theta_i + \rho\phi_i)^3}} . \qquad (3.3.2)$$

If the matching is performed, the expected cell frequencies will correspond to those of Table 1.2, and the expected matched X^2 statistic is:

$$X^2 = X^2_m$$

$$= \frac{\left[\sum_i \frac{(n_1\alpha_i\theta_i)(n_0(1-\beta_i)\theta_i) - (n_1(1-\alpha_i)\theta_i)(n_0\beta_i\theta_i)}{n_1\theta_i + n_0\theta_i}\right]^2}{\sum_i \frac{(n_1\alpha_i\theta_i + n_0\beta_i\theta_i)[(n_1\theta_i + n_0\theta_i) - (n_1\alpha_i\theta_i + n_0\beta_i\theta_i)]n_1n_0\theta_i^2}{(n_1\theta_i + n_0\theta_i)^3}}$$

$$= \frac{n_1\rho(1+\rho)(RR-1)^2\left[\sum_i \theta_i\beta_i\right]^2}{\sum_i \theta_i\beta_i(RR+\rho)[(1+\rho)-\beta_i(RR+\rho)]} . \qquad (3.3.3)$$

For the case of matching, the conditions regarding uniformity and the sampling ratio are also incorporated. In addition, the consideration of loss of subjects due to the matching process is an important issue regarding efficiency that has often been ignored. The loss can be expressed in terms of the proportion of subjects which remain after the matching has been performed. Allowing different proportions of loss for the index sample and the reference sample, let

r_1 = proportion of index sample which is matced,

r_0 = proportion of referent sample which is matched.

In (3.3.3) replace n_1 with $r_1 n_1$ and n_0 with $r_0 \rho n$, as well as α_i with $RR\beta_i$. The expected matched X^2 statistic is then expressed as

$$X_m^2 = \frac{n_1(r_1+r_0\rho)r_1r_0\rho(RR-1)^2\left[\sum_i \theta_i\beta_i\right]^2}{\sum_i \theta_i\beta_i(r_1RR+r_0\rho)[(r_1+r_0\rho)-\beta_i(r_1RR+r_0\rho)]} . \qquad (3.3.4)$$

Let the relative efficiency (RE) be the ratio X_m^2 to X_{rs}^2. The expression for RE is then given by:

$$RE = \frac{(r_1+r_0\rho)r_1r_0\left[\sum_i \theta_i\beta_i\right]^2\left[\sum_i \frac{\theta_i\phi_i\beta_i(RR\theta_i+\rho\phi_i)[(\theta_i+\rho\phi_i)-\beta_i(RR\theta_i+\rho\phi_i)]}{(\theta_i+\rho\phi_i)^3}\right]}{\left[\sum_i \theta_i\phi_i\beta_i/(\theta_i+\rho\phi_i)\right]^2\left[\sum_i \theta_i\beta_i(r_1RR+r_0\rho)[(r_1+r_0\rho) - \beta_i(r_1RR+r_0\rho)]\right]} .$$

(3.3.5)

This index is asymptotically equivalent to the Pitman efficiency for alternatives near the null (see Appendix 3).

3.3.2 Case-Control Study

For the case-control study, similar expressions are derived. The uniformity assumption is given by:

$$OR = \frac{\varepsilon_i(1-\delta_i)}{\delta_i(1-\varepsilon_i)} \Rightarrow \varepsilon_i = \frac{\delta_i OR}{(1-\delta_i(1-OR))} , \quad \forall_i .$$

The parameters θ_i, ϕ_i, β_i, and RR are replaced by v_i, w_i, δ_i, and OR respectively, and (3.3.2), (3.3.4), and (3.3.5) are expressed in analogous fashion as

$$\chi^2_{rs} = \frac{\rho n_1 \left[\sum_i v_i w_i \delta_i (f_i-1)/(v_i+\rho w_i)\right]^2}{\sum_i \frac{v_i w_i \delta_i (f_i+\rho w_i)[(v_i+\rho w_i)-\delta_i(f_i+\rho w_i)]}{(v_i+\rho w_i)^3}} \tag{3.3.6}$$

$$\chi^2_m = \frac{n_1(r_1+r_0\rho)r_1r_0\rho\left[\sum_i v_i \delta_i (f_i-1)\right]^2}{\sum_i v_i\delta_i(r_1f_i+r_0\rho)[(r_1+r_0\rho)-\delta_i(r_1f_i+r_0\rho)]} \tag{3.3.7}$$

where

$$f_i = OR \cdot (1-\delta_i(1-OR))^{-1} .$$

In the same way the relative efficiency is expressed as the ratio of (3.3.7) to (3.3.6).

3.4 Results of Deterministic Evaluations

3.4.1 No Loss from the Matching

For both of the studies, suitable ranges of the values of the parameters were chosen and substituted in the expressions for RE. In keeping with the properties of the Pitman efficiency, the values chosen to represent the measure of effect (RR and OR) are both in the neighborhood of the null (1.5 and 2.25). For these analyses, only one dichotomous (potentially) confounding variable was considered. Nine distributions of this variable conditioning on index or referent status were included in the analyses. These distributions are represented by $\theta_1(v_1) = .1,.2,\ldots,.9$ and $\phi_1(w_1) = .1,.2,\ldots,.9$.

Additional parameters which were varied in the evaluations represent the degree of confounding in the population, the level of the overall disease or exposure probabilities (represented by β_0 and δ_0, respectively), and the referent/index sampling ratio, ρ. In this section the loss of sample size due to matching will be assumed to be zero. The results of evaluations based upon non-zero loss are presented in Section 3.4.3.

Some general conclusions can be reached from these evaluations. First, matching will on the average lead to an expected gain in efficiency over random sampling. Averaging over all levels of confounding in the population and distributions of the extraneous variable, as well as considering a range of values for the other parameters, the (geometric) mean gain in efficiency is roughly 15 to 20 percent. This result is not restricted to either type of study or measure of effect.

Tables 3.5 through 3.8 summarize these conclusions in the column headed by "Overall Gain." Average gains are calculated by the following formula, representing the follow-up study,

$$\text{GAIN} = -1 + \left[\prod_{\theta_1=.1}^{.9} \prod_{\phi_1=.1}^{.9} \text{RE}(\theta_1,\phi_1) \right]^{1/81} .$$

Other average gains or losses reported in this chapter are calculated in a similar fashion.

Second, there is an important relationship between relative efficiency and the nature of confounding. Like any other bias, confounding bias may be positive or negative; that is, the bias may result in a measure of effect which is pushed upward or pulled downward away from the true value. To distinguish between these two types of bias they will be referred to as "positive" and "negative" confounding. In terms of the population parameters, positive and negative confounding are characterized by particular values of the odds ratios which describe the association between F, the extraneous variable, and exposure and disease, as defined in Chapter 1. Table 3.1 summarizes the correspondence between these associations and the types of confounding.

The expected relative efficiency is markedly affected by the type of confounding in the population. (Recall that the control of confounding is presumed in these evaluations. The reference to confounding concerns those relationships among the variables which would lead to confounding, were it not controlled for.) When there is positive confounding in the population, matching is universally more efficient than random sampling. This expected gain in efficiency from matching is not insignificant and can be quite substantial, when there is

TABLE 3.1

Confounding and Associations
Between F, Disease, and Exposure

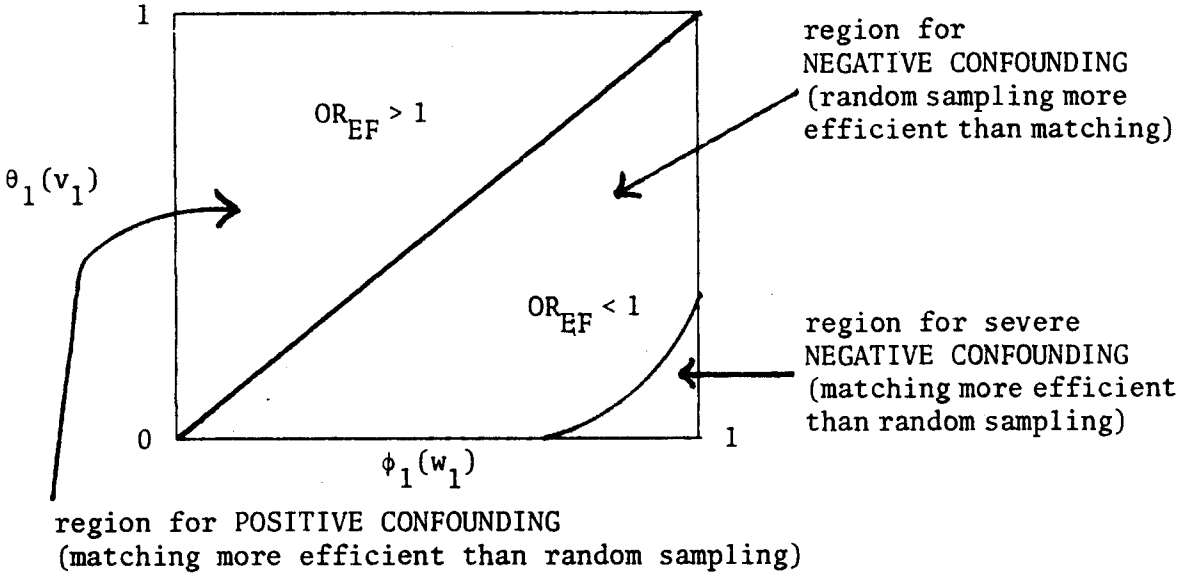
Type of Confounding	Associations	
	Follow-up	Case-Control
Positive	$OR_{EF} > 1$ and $OR_{DF/\bar{E}} > 1$	$OR_{EF/\bar{D}} > 1$ and $OR_{DF/\bar{G}} > 1$
	or	or
	$OR_{EF} < 1$ and $OR_{DF/\bar{E}} < 1$	$OR_{EF/\bar{D}} < 1$ and $OR_{DF/\bar{E}} < 1$
Negative	$OR_{EF} > 1$ and $OR_{DF/\bar{E}} < 1$	$OR_{EF/\bar{D}} > 1$ and $OR_{DF/\bar{E}} < 1$
	or	or
	$OR_{EF} < 1$ and $OR_{DF/\bar{E}} > 1$	$OR_{EF/\bar{D}} < 1$ and $OR_{DF/\bar{E}} > 1$

severe positive confounding. On the other hand, when there is negative confounding in the population, random sampling is generally more efficient, although to a lesser magnitude. In some cases, if there is severe negative confounding in the population, matching can lead to a greater expected efficiency than random sampling. These distinctions were noted by Kupper, et al. (1980) and are portrayed in Figure 3.1.

A third conclusion can be made regarding the referent/index sample size ratio, ρ. Contrasting Tables 3.5 and 3.7 to Tables 3.6 and 3.8, respectively, it is evident that increasing this ratio causes the distinctions between matching and random sampling to be less pronounced. In other words, the advantages of matching are more subdued but so are the disadvantages. The effect of increasing the sampling ratio is therefore dependent on the direction of confounding. While matching usually leads to an expected loss in efficiency when there is negative

FIGURE 3.1

Regions of Confounding[1]



[1]Note the assumption that $OR_{DF/\overline{E}} > 1$. Also, for a case-control study OR_{EF} is replaced with OR_{DF} and $OR_{DF/\overline{E}}$ with $OR_{EF/\overline{D}}$.

confounding, increasing ρ from one to three in the context of a case-control study can turn that expected loss into an expected gain, or at least a negligible expected loss (see Tables 3.7 and 3.8, column 5).

Fourth, relative efficiency is not a function of the rarity of the disease in a follow-up study or the probability of exposure in a case-control study. Although the decision to match may be dependent on the probability of disease (or exposure), due to small expected cell frequencies, if matching can indeed be accomplished, the expected gain over random sampling is quite unrelated to this circumstance. These results will now be considered in more detail for the follow-up and case-control studies.

3.4.2 Follow-up Study

Deterministic analyses were performed after specifying the parameters listed in Table 3.2. The values chosen for these parameters were considered representative of practical situations in which follow-up studies would be conducted.

TABLE 3.2

Follow-up Study Parameters

Parameter	Values
RR	1.5, 2.25
$\beta_0 = P(D/EF_o)$	0.0001, 0.01
$OR_{DF/\overline{E}} = \frac{\beta_1(1-\beta_0)}{\beta_0(1-\beta_1)}$	2.0, 5.0
$OR_{EF} = \frac{\theta_1\phi_0}{\theta_0\phi_1}$	(see Table 3.2)
$\rho = n_1/n_0$	1 and 3

The summaries in Tables 3.5 and 3.6 average over the OR_{EF} values. More complete results are given in Tables 3.9 to 3.11, which indicate that the relative efficiency is a complex function of the parameters listed above. The relationship between the relative efficiency and each parameter will be considered separately.

Clearly, the most important determinant of RE is the degree of confounding, as measured by OR_{EF} and $OR_{DF/\overline{E}}$. Each of Tables 3.9 to 3.11 displays RE for values of OR_{EF} ranging from 1/81 to 81, as

well as the values 2.0 and 5.0 for $OR_{DF/\bar{E}}$. Table 3.3 lists each of the values of OR_{EF} considered. (See also Figure 1. Note that these values are sufficient to characterize conditions for both positive and negative confounding.) When OR_{EF} is moderate to high in value, RE ranges from 1.5 to 2.0 and from 3.9 to 5.1 for more extreme values. In addition, RE changes by as much as 10% to 25% by increasing $OR_{DF/\bar{E}}$ from 2 to 5 (Tables 3.9 and 3.10).

TABLE 3.3

Values Specified for OR_{EF} in a Follow-up Study and OR_{DF} in a Case-Control Study

		ϕ_1				
		0.1	0.3	0.5	0.7	0.9
θ_1	0.9	81.0	21.0	9.0	3.86	1.0
	0.7	21.0	5.44	2.33	1.0	0.26
	0.5	9.0	2.33	1.0	0.43	0.11
	0.3	3.86	1.0	0.43	0.18	0.05
	0.1	1.0	0.26	0.11	0.05	0.01

Varying the levels of β_0 and ρ, in general, yields results which correspond to those reported earlier for both studies. If the risks of disease increase and ρ is increased, RE decreases when there is positive confounding and increases when there is negative confounding. Under the conditions of severe confounding, the impact of increasing ρ on RE can be substantial. It should be noted that most of the influence on RE realized by increasing ρ occurs from the increase of $\rho = 1$ to 2.

3.4.3 Case-Control Study

The overall results of the evaluations for this design are quite similar to those of the follow-up study. Actually RE is a function of identical parameters, with the exception of the measure of effect (odds ratio) and the measure of the exposure/confounder association ($OR_{EF/\bar{D}}$). Table 3.4 lists the parameters which were specified along with the values that were considered

TABLE 3.4

Case-Control Study Parameters[1]

Parameter	Values
OR	1.5, 2.25
$\delta_0 = P(E/\bar{D}F_0)$	0.01, 0.05
$OR_{EF/\bar{D}} = \frac{\delta_1(1-\delta_0)}{\delta_0(1-\delta_1)}$	2.0, 5.0
$OR_{DF} = \frac{v_1 w_0}{v_0 w_1}$	(see Table 3.2)
$\rho = n_1/n_0$	1 and 3

1. The parameter used to measure the association between F and disease in a case-control study is

$$OR_{DF/\bar{E}} = \frac{(1-\varepsilon_1)v_1(1-\delta_0)w_0}{(1-\varepsilon_0)v_0(1-\delta_1)w_1} = OR_{DF} \cdot \frac{(1-\delta_0(1-OR))}{(1-\delta_1(1-OR))}$$

The primary difference in these specifications relative to those of the follow-up study are the levels of the exposure probabilities,

represented by δ_0. In practical situations, we should expect exposure probabilities conditional on disease status to be much higher than conditional probabilities of disease in follow-up studies.

Tables 3.7 and 3.8 summarize the results of the evaluations. Matching provides a 15% to 25% average expected gain over random sampling and a 25% to 65% average gain when there is positive confounding in the population. Losses from matching under the conditions of negative confounding range up to 15%. Compared to the follow-up study, matching seems more likely to result in a gain in efficiency over random sampling although to a lesser extent.

Tables 3.12 and 3.13 give examples of RE for various values of the parameters listed in Table 3.4. Most of the discussion in Section 3.4.1a pertains to the case-control study as well. A notable exception is the parameter δ_0. Under strong confounding, where $OR_{EF/\overline{D}} = 5$, if δ_0 is relatively large, then the RE is less extreme than is otherwise the case. That is, when the exposure probabilities are uniformly large, the differences in efficiency between matching and random sampling are less distinctive.

In summary, matching is a more efficient method of subject selection in a case-control study when the underlying confounding is positive ($OR_{DF/\overline{E}} > 1$) or severely negative ($OR_{DF/\overline{E}} << 1$), when the exposure probabilities are small, and when the sampling ratio, ρ, is 1. Random sampling is more efficient when there is moderate negative confounding and the sampling ratio and exposure probabilities are relatively large.

3.4.4 Loss Due to Matching

Up until now the results of the relative efficiency evaluations have presumed that no loss of subjects from the matching process occurred. For typical epidemiologic studies, however, there is indeed a loss of subjects from the matching. McKinlay (1974) has documented studies in which matching was performed and the losses which resulted ranged up to 80%. Since a loss in sample size from matching seems to be certain, the question which arises is: how much is the loss and how can it be measured?

Frequency-matching can actually be accomplished in three different ways: (a) adjusting the index series to correspond to the distribution of the matching variables in the referent series; (b) adjusting the referent series to correspond to the index series; and (c) adjusting both series in such a way that there is correspondence with regard to the variables of interest. The first method is rarely used since the index series is usually small and expensive to construct. The second method (Method B) is more likely to be used when there is a large pool of referents available and large losses can be absorbed. The advantage of this method is that the distribution of the matching variables in the index series is maintained. The third method (Method C) is the usual method when sample sizes are relatively small, although greater losses of sample size can occur compared to Method B. In general, however, this method yields smaller losses in efficiency.

Very little work has been done on the quantification of loss due to matching. McKinlay (1974) suggested a mathematical model which involves the assumption of probability distributions on the matching

categories of each population. The model leads to estimates of the expected number of matches and the variance given fixed sample sizes and a 1:1 matching ratio. Implied in this model is matching according to Method C. Chase (1968) also derived a model for the number of matches but a critical assumption is necessary regarding the distribution of the outcome variable, rendering his approach less useful.

In this context the quantification of loss is not alone of primary concern but rather is of secondary concern to how loss relates to relative efficiency. Therefore, the distributional approach taken by McKinlay is not necessary in order to provide adequate measures of loss. Instead, it is sufficient to write the loss in terms of the population parameters describing the joint distributions of the matching variables. This approach will be accomplished for Methods B and C.

3.4.5 Quantification of Loss

Let us assume with no loss in generality that in a follow-up study, n_1 exposed subjects and ρn_1 unexposed subjects will be selected by random sampling and then matched. After random sampling the expected number of exposed and unexposed subjects in the i^{th} stratum of F is $n_1\theta_i$ and $\rho n_1 \phi_i$, respectively. If Method B matching is performed, the unexposed series will be adjusted to correspond in distribution of F to the exposed series. Hence, the ratio of the number of exposed to unexposed subjects will be constant across the strata of F. What should the value of this ratio be? For the i^{th} stratum of F the ratio of unexposed subjects to exposed before matching is

$$\frac{\rho n_1 \phi_i}{n_1 \theta_i} = \frac{\rho \phi_i}{\theta_i} .$$

To adjust the unexposed to correspond properly to the exposed series without sampling additional subjects, the ratio adopted must be the minimum across the strata. Therefore, unexposed subjects from some strata will be discarded to bring the ratio in that stratum to the minimum, which occurs at $\min_i(\phi_i/\theta_i)$. The proportion of referents which are matched is then given by

$$r_0 = \min_i \frac{\phi_i}{\theta_i} . \qquad (3.4.1)$$

Since there is no loss of exposed subjects, $r_1 = 1$. Note that $r_0 < 1$ unless $\theta_i = \phi_i$, $\forall_i$.

Under matching by Method C both the exposed and unexposed series will be adjusted to some artificial distribution across F. The matching process compares the number of exposed subjects and unexposed subjects in each stratum, and discards subjects from the larger cell until there are equal numbers in the two cells. The number of exposed (and unexposed) subjects which are matched in the i^{th} stratum is

$$\min(n_1\theta_i, \rho n_1 \phi_i) = n_1 \cdot \min(\theta_i, \rho\phi_i) .$$

Therefore, the total number of exposed (and unexposed) subjects matched are

$$n_1 \cdot \sum_i \min(\theta_i, \rho\phi_i) .$$

The proportions of exposed and unexposed subjects which are matched are given by

$$r_1 = \frac{n_1 \sum_i \min(\theta_i, \rho\phi_i)}{n_1} = \sum_i \min(\theta_i, \rho\phi_i) \quad (3.4.2)$$

$$r_0 = \frac{n_1 \sum_i \min(\theta_i, \rho\phi_i)}{\rho n_1} = \frac{\sum_i \min(\theta_i, \rho\phi_i)}{\rho} . \quad (3.3.4)$$

As ρ grows large, $r_1 \to 1$ and $r_0 \to 1/\rho$. Hence, Method C matching approximates Method B matching under the condition of a large sampling ratio, but with the restriction that the matched samples are the same size. Since the loss of referents is large under these conditions, it is preferable to use Method B, which allows for a larger number of referents to be matched. Under Method C matching, the distribution of the matching variable, $\underset{\sim}{\theta}$, is not maintained by the matching. Rather, a new distribution, $\underset{\sim}{\Delta}$, is created which does not correspond to $\underset{\sim}{\theta}$ nor to $\underset{\sim}{\phi}$, the distribution of the matching variables in the unexposed. This new distribution is given by

$$\Delta_i = \frac{\min(\theta_i, \rho\phi_i)}{\sum_i \min(\theta_i, \rho\phi_i)} .$$

An important implication of these methods of quantifying loss is that the loss of sample size increases with the disparity between θ_i and ϕ_i. In Section 3.4.1 where no loss of sample size was presumed, matching was shown to be most advantageous precisely under these conditions. A natural question to ask is whether the advantages of matching under no loss are overcome by the substantial loss of sample size experienced under the conditions described. In the following section results of deterministic evaluations are presented which address this question.

3.4.6 Loss of Sample Size and Relative Efficiency

Evaluations of RE under both Method B and Method C quantification of loss were conducted substituting the same values of the parameters as previously (assuming no loss). Some general conclusions can be drawn from this work:

(a) Regardless of the choice of matching method (quantification of loss), if there is confounding, matching is without exception less efficient than random sampling. The average and individual gains expected from matching, reported in Section 3.4.1 are universally reversed when the loss involved in the matching process is incorporated.

(b) The loss in efficiency from matching increases as confounding increases despite the fact that the gain in efficiency from matching (assuming no loss in sample size) increases with increasing confounding.

(c) Method B loss in sample size is less than Method C loss; however, the loss in efficiency is generally greater for Method B matching (50% on the average) if the sampling ratio is 1. This disparity almost disappears when the ratio is 3, as the two methods of matching are essentially identical.

An example of the loss in efficiency which results from matching when there is a loss in sample size is given in Table 3.14. The

expected RE is given for a case-control study where OR = 1.5, δ_0 = 0.05, $OR_{EF/\bar{D}}$ = 2, ρ = 1, and Method B matching is presumed. Corresponding to these results are those of Table 3.12, where the same values of the parameters are used, but no loss is presumed. A comparison of the two tables reveals the drastic effect of the loss in sample size on the relative efficiency.

If no loss in sample size is presumed, matching gains an average of 18.4%. But when loss is incorporated, matching averages a 26.5% loss in efficiency to random sampling. Note that the reversal is most extreme where the gain from matching was greatest, at the outer corners of the table where confounding is strong. This comparison typifies the general result of incorporating loss of subjects in the study of relative efficiency: that matching will not compare favorably to random sampling.

3.5 Recommendations

Based on the results summarized in this chapter, the choice to match or to forego matching should be easily made. Generally, regardless of the type of study design, matching should be employed if no loss of sample size will occur. The typical gain in efficiency which may obtain from the matched design and analysis is worth the risk of a loss in efficiency. If one can determine a priori the nature of the confounding (via specific associations involving F, the exposure, and the disease), then the decision to match is even more clearcut.

On the other hand, if one can expect to lose sample size from the matching, then matching should not be employed as the method of subject selection.

These recommendations are made with the qualification that only one variable measured at two levels is the basis of the matching. Additional matching variables involving increased strata may modify the considerations involved in choosing one method of subject selection over the other. The evaluation of relative efficiency in the context of multiple matching variables is the subject of Chapter 5.

TABLE 3.5

Average Gains in Efficiency from Matching: Follow-up Study, $\rho = 1$

RR	β_0	Overall	Among Gains	Among Losses	Positive Confounding	Negative Confounding
			$OR(DF/\bar{E}) = 2.0$			
1.5	0.0001	18.5%	32.4%	- 2.7%	38.4%	5.9%
	0.01	18.5	32.4	- 2.7	38.5	5.8
2.25	0.0001	18.4	35.5	- 3.6	40.9	3.8
	0.01	18.3	35.3	- 3.6	40.8	3.7
			$OR(DF/\bar{E}) = 5.0$			
1.5	0.001	16.4%	54.7%	-11.9%	57.6%	-10.7%
	0.01	16.3	54.7	-12.1	57.7	-11.0
2.25	0.001	16.0	61.6	-15.0	63.6	-14.6
	0.01	15.8	60.7	-14.9	62.7	-14.5

TABLE 3.6

Average Gains in Efficiency from Matching:
Follow-up Study, $\rho = 3$

RR	β_0	Overall	Among Gains	Among Losses	Positive Confounding	Negative Confounding
			$OR(DF/\overline{E}) = 2.0$			
1.5	0.001	17.3%	24.2%	-0.9%	29.3%	10.7%
	0.01	17.2	24.1	-0.9	29.2	10.6
2.25	0.0001	20.2	29.0	-1.3	35.4	11.6
	0.01	20.0	28.8	-1.2	35.0	11.6
			$OR(DF/\overline{E}) = 5.0$			
1.5	0.0001	15.3%	33.8%	-4.4%	38.7%	-0.6%
	0.01	15.1	33.4	-4.4	38.3	-0.7
2.25	0.0001	17.6	41.7	-6.1	47.5	-2.3
	0.01	17.3	40.6	-5.9	46.2	-2.1

TABLE 3.7

Average Gains in Efficiency from Matching:
Case-Control Study, $\rho = 1$

OR	δ_0	Overall	Among Gains	Among Losses	Positive Confounding	Negative Confounding
			$OR(EF/\bar{D}) = 2.0$			
1.5	0.01	18.5%	31.6%	-2.5%	38.5%	6.3%
	0.05	18.5	29.9	-1.9	35.5	8.3
2.25	0.01	18.4	34.3	-3.2	39.8	4.6
	0.05	18.4	29.7	-2.1	35.5	7.9
			$OR(EF/\bar{D}) = 5.0$			
1.5	0.01	16.6%	51.6%	-10.9%	55.6%	- 9.3%
	0.05	17.2	41.4	- 6.7	47.2	- 2.9
2.5	0.01	16.3	56.2	-13.0	59.4	-11.9
	0.05	17.1	39.9	- 5.6	44.4	- 1.3

TABLE 3.8

Average Gains in Efficiency from Matching:
Case-Control Study, $\rho = 3$

OR	δ_0	Overall	Among Gains	Among Losses	Positive Confounding	Negative Confounding
			$OR(EF/\bar{D}) = 2.0$			
1.5	0.01	17.2%	24.1%	-0.8%	28.8%	10.9%
	0.05	16.8	27.3	-0.6	26.8	11.9
2.25	0.01	19.9	28.1	-1.1	34.2	12.1
	0.05	18.9	24.2	-0.7	29.8	13.8
			$OR(EF/\bar{D}) = 5.0$			
1.5	0.01	15.3%	32.2%	-3.8%	37.1%	0.5%
	0.05	15.1	25.5	-2.0	30.8	4.9
2.25	0.01	17.4	37.6	-4.9	43.8	-0.2
	0.05	16.5	25.8	-1.5	30.9	7.7

TABLE 3.9

Relative Efficiency of Matching and Random Sampling: Follow-up Study

($RR = 1.5$, $\beta_0 = 0.0001$, $\rho = 1$, $OR_{DF/\overline{E}} = 2$)

		ϕ_1				
		0.1	0.3	0.5	0.7	0.9
θ_1	0.9	3.71	1.87	1.35	1.12	1.0
	0.7	2.15	1.39	1.12	1.0	0.97
	0.5	1.55	1.14	1.0	0.96	~1.0
	0.3	1.21	1.0	0.96	~1.0	1.14
	0.1	1.0	~1.0	1.12	1.36	1.93

Mean Gain (Positive Conf.) = 38.4%

Mean Gain (Negative Conf.) = 5.9%

Mean Overall Gain = 18.5%

TABLE 3.10

Relative Efficiency of Matching and Random Sampling: Follow-up Study

$(RR = 2.25\ ,\ \beta_0 = 0.01\ ,\ \rho = 1\ ,\ OR_{DF/\overline{E}} = 5)$

		ϕ_1				
		0.1	0.3	0.5	0.7	0.9
θ_1	0.9	5.04	2.16	1.47	1.17	1.0
	0.7	3.21	1.65	1.20	1.0	0.89
	0.5	2.23	1.29	1.0	0.87	0.81
	0.3	1.53	1.0	0.84	0.79	0.76
	0.1	1.0	0.85	0.86	0.92	1.02

Mean Gain (Positive Conf.) = 62.7%

Mean Gain (Negative Conf.) = -14.5%

Mean Overall Gain = 15.8%

TABLE 3.11

Relative Efficiency of Marching and Random Sampling: Follow-up Study

($RR = 2.25$, $\beta_0 = 0.01$, $\rho = 3$, $OR_{DF/\bar{E}} = 5$)

		ϕ_1				
		0.1	0.3	0.5	0.7	0.9
θ_1	0.9	3.90	1.77	1.31	1.11	1.0
	0.7	2.77	1.45	1.13	1.0	0.95
	0.5	1.99	1.20	1.0	0.93	0.93
	0.3	1.40	1.0	0.92	0.92	0.99
	0.1	1.0	0.94	~1.0	1.15	1.50

Mean Gain (Positive Conf.) = 46.2%

Mean Gain (Negative Conf.) = -2.1%

Mean Overall Gain = 17.3%

TABLE 3.12

Relative Efficiency of Matching and Random Sampling:
Case-Control Study

$(OR = 1.5\ ,\ \delta_0 \doteq 0.05\ ,\ \rho = 1\ ,\ OR_{EF/\overline{D}} = 2)$

		w_1				
		0.1	0.3	0.5	0.7	0.9
v_1	0.9	3.84	1.96	1.39	1.13	1.0
	0.7	2.15	1.41	1.13	1.0	0.96
	0.5	1.55	1.14	1.0	0.95	0.97
	0.3	1.21	1.0	0.95	0.98	1.08
	0.1	1.0	~1.0	1.11	1.36	1.85

Mean Gain (Positive Conf.) = 35.5%

Mean Gain (Negative Conf.) = 8.3%

Mean Overall Gain = 18.5%

TABLE 3.13

Relative Efficiency of Matching and Random Sampling: Case-Control Study

$(OR = 2.25\ ,\ \delta_0 = 0.01\ ,\ \rho = 1\ ,\ OR_{EF/\overline{D}} = 5)$

		w_1				
		0.1	0.3	0.5	0.7	0.9
v_1	0.9	4.86	2.13	1.46	1.16	1.0
	0.7	3.05	1.62	1.19	1.0	0.90
	0.5	2.12	1.27	1.0	0.88	0.83
	0.3	1.49	1.0	0.86	0.81	0.80
	0.1	1.0	0.87	0.90	0.98	1.12

Mean Gain (Positive Conf.) = 59.4%

Mean Gain (Negative Conf.) = -11.9%

Mean Overall Gain = 16.3%

TABLE 3.14

Relative Efficiency of Matching and Random Sampling: Case-Control Study,

Loss of Sample Size from Matching: Method B

($OR = 1.5$, $\delta_0 = 0.05$, $\rho = 1$, $OR_{EF/\overline{D}} = 2$)

		w_1				
		0.1	0.3	0.5	0.7	0.9
v_1	0.9	0.57	0.68	0.77	0.88	1.0
	0.7	0.62	0.72	0.85	1.0	0.86
	0.5	0.72	0.83	1.0	0.84	0.74
	0.3	0.83	1.0	0.82	0.70	0.62
	0.1	1.0	0.83	0.72	0.63	0.54

Mean Gain (Positive Conf.) = -26.0%

Mean Gain (Negative Conf.) = -27.8%

Mean Overall Gain = -26.5%

CHAPTER 4

VALIDITY AND MULTIPLE CONFOUNDING VARIABLES

4.1 Introduction

A large portion of the methodologic research on confounding and matching in epidemiologic studies has concerned only the most elementary situation, that which involves a single, potentially confounding variable. Even so, the limits to which this situation can be researched have not yet been reached. And, despite the elementary nature of the case of one confounding variable, analytical solutions to many of the problems under study do not exist. For these reasons, among others, the consideration of multiple confounding variables has mostly been ignored or limited to numerical studies.

However, the control of multiple confounding variables is an important issue since most epidemiologic studies require the control of from two to five potentially confounding variables (Billewicz, 1964). While the research concerning a single confounder is not irrelevent to the question of multiple confounders, Miettinen (1974) has correctly warned that the failure to find confounding or effect modification from the study of the marginal relationships alone can mask the presence of either or both phenomenon, when the joint relationships are considered. Kleinbaum, et al. (1979) show by example

the results of the failure to adjust for multiple variables simultaneously, as well as the complexities involved in doing so. In addition, Fisher and Patil (1974) attempt to convey this principle through some artificial numerical examples.

The topic of validity in the context of a single confounding variable has been quite thoroughly researched by Miettinen (1974) and Kupper, et al. (1979). In this chapter the characterization of confounding is extended to multiple (potentially) confounding variables. In particular, conditions for no confounding are developed involving two extraneous variables. Two approaches are utilized: 1) comparisons of crude and adjusted measures of effect, and 2) a regression formulation of risk as a function of the extraneous variables and exposure. These investigations suggest that the accepted definition of a confounding variable (Chapter 1) breaks down when there are multiple extraneous variables considered simultaneously.

The implications of matching on the control of confounding, especially with respect to case-control studies are also considered in this chapter. There remains controversy over the proper analysis of matched data, for validity purposes, in a case-control study (see Kupper, et al., 1980). The results in this chapter show that, in practice, stratification need not follow matching when the matching variables include all of the confounding variables. If there are confounders which have not been matched on, however, stratification will usually be required on the matching variables as well as the other confounders.

Finally, the mixture of the effects of confounding and effect modification are investigated in this chapter. The apparent presence

of effect modification, when there is in fact uniformity of effect, can result from the failure to control for a confounding variable. An example of this phenomenon is presented.

4.2 Validity and Multiple Confounding Variables

The concept of confounding grows more elusive as additional variables are introduced. Before studying the nature of confounding in the context of multiple variables, it would be helpful to develop some definitions.

First, if there is confounding due to two or more variables simultaneously, this will be called joint confounding. Marginal confounding will be described by confounding due to a single variable based upon the marginal distributions of that variable, disease, and exposure. The conditions for no confounding outlined by (1.3.2a) and (1.3.2b) are those for no marginal confounding. A third type of confounding, which will be considered in more detail in a later section, is residual confounding. An apt description of this type of confounding is the "left-over" confounding due to a variable G after adjusting for other variables.

4.2.1 Follow-up Study: Conditions for No Confounding

Consider a study which involves k (potentially) confounding variables. The follow-up parameters defined in Chapter 1 represent this extension, where $i = 1,\ldots,L$ and L denotes the total number of joint levels of the k confounders. The stratum-specific and expected cell frequencies do not change from those in Table 1.1.

The conditions for no joint confounding in this context can be derived directly from (1.3.2.3). If the joint distribution of the k confounders among the exposed (θ, i = 1,...,L) is the standard, then there is no confounding if:

$$cRR = sRR \Leftrightarrow \frac{\sum_i^L \alpha_i\theta_i}{\sum_i^L \beta_i\phi_i} = \frac{\sum_i^L \alpha_i\theta_i}{\sum_i^L \beta_i\theta_i}, \quad \text{or}$$

$$\sum_i^L \beta_i\phi_i = \sum_i^L \beta_i\theta_i ,$$

which can be expressed as

$$\sum_i^{L-1} \beta_i(\theta_i - \phi_i) + \beta_L\left[1 - \sum_i^{L-1} \theta_i - \left(1 - \sum_i^{L-1} \phi_i\right)\right] = 0 , \quad \text{or}$$

$$\sum_i^{L-1} (\beta_i - \beta_L)(\theta_i - \phi_i) = 0 . \qquad (4.2.1)$$

The number of solutions to (4.2.1) is infinite, however, there are a few which are of special interest. There is no confounding if either:

$$(a) \quad \beta_i = \beta_L , \quad \forall\ i = 1,...,L-1$$

or

$$(b) \quad \theta_i = \phi_i , \quad \forall\ i = 1,...,L-1 . \qquad (4.2.2)$$

These conditions are direct extensions of the conditions for no confounding for the single confounder case. Condition (a) can be interpreted as "the risk of disease among all subcategories of the unexposed is uniform." Condition (b) implies that the distribution of F is independent of exposure status.

By introducing additional levels to the confounding variable (or, equivalently, additional variables) the conditions outlined by (4.2.2) are now sufficient but not necessary for no confounding. An endless variety of conditions that meet (4.2.1) could be imagined. From the point of view of controlling confounding, however, the lack of the necessity of (4.2.2) is unimportant. Matching on F (or the joint distribution of multiple extraneous variables) will indeed control confounding in a follow-up study by meeting (4.2.2b). Since the matching replaces the ϕ_i's by the θ_i's in the analysis, a crude analysis yields the sRR.

4.2.2 Case-Control Study: Conditions for No Confounding

Extending the conditions for no confounding from one dichotomous confounder to L levels proceeds in the same manner as the follow-up study. For the case of L categories, there is no confounding if:

$$cOR = sOR \Longleftrightarrow \frac{\left[\sum_i \varepsilon_i v_i\right]\left[\sum_i (1-\delta_i) w_i\right]}{\left[\sum_i \delta_i w_i\right]\left[\sum_i (1-\varepsilon_i) v_i\right]} = \frac{\left[\sum_i \varepsilon_i v_i\right]}{\sum_i \frac{(1-\varepsilon_i) v_i \delta_i}{1-\delta_i}}, \quad \text{or}$$

$$\sum_i \frac{(1-\varepsilon_i) v_i \delta_i}{1-\delta_i} = \frac{\left[\sum_i \delta_i w_i\right]\left[\sum_i (1-\varepsilon_i) v_i\right]}{\sum_i (1-\delta_i) w_i}, \quad \text{or}$$

$$\left[\sum_i (1-\delta_i) w_i\right]\left[\sum_i \frac{(1-\varepsilon_i) v_i \delta_i}{(1-\delta_i)}\right] = \left[\sum_i \delta_i w_i\right]\left[\sum_i (1-\varepsilon_i) v_i\right], \quad \text{or}$$

$$\sum_i \frac{(1-\varepsilon_i) v_i \delta_i}{1-\delta_i} - \left[\sum_i \delta_i w_i\right]\left[\sum_i \frac{(1-\varepsilon_i) v_i \delta_i}{1-\delta_i}\right] = \left[\sum_i \delta_i w_i\right]\left[\sum_i (1-\varepsilon_i) v_i\right], \quad \text{or}$$

$$\sum_i \frac{(1-\varepsilon_i)v_i\delta_i}{1-\delta_i} = \left(\sum_i \delta_i w_i\right)\left[\sum_i\left\{(1-\varepsilon_i)v_i + \frac{(1-\varepsilon_i)v_i}{(1-\delta_i)}\right\}\right]$$

$$= \left(\sum_i \delta_i w_i\right)\left(\sum_i \frac{(1-\varepsilon_i)v_i}{1-\delta_i}\right) .$$

Therefore,

$$\sum_i \frac{(1-\varepsilon_i)v_i}{1-\delta_i}\left(\delta_i - \sum_i \delta_i w_i\right) = 0 . \qquad (4.2.3)$$

We can identify two conditions from this expression which will imply no confounding:

(a) $\delta_i = \delta$, $\forall\ i = 1,\ldots,L$

or

(b) $\dfrac{(1-\varepsilon_i)v_i}{(1-\delta_i)w_i} = c$, a constant, $\forall\ i = 1,\ldots,L$. (4.2.4)

Note that these conditions carry over directly from the conditions for no confounding in the single confounder setting (Chapter 1). Condition (a) is interpreted as "the probability of exposure among all subcategories of the controls is uniform." Condition (b) implies that "the ratio of the joint probability of non-exposure and the i^{th} level of F conditional on disease to the joint probability of the same event conditional on non-disease is uniform across all strata of F." In other words condition (b) states that the joint distributions of exposure and F conditional on case status and control status are identical. These conditions will be considered in more detail for the case of two dichotomous confounding variables.

Again (4.2.4a) and (4.2.4b) are not necessary for (4.2.3) to hold. However, as discussed in the context of the follow-up study, such non-necessity is irrelevent from the point of view of controlling confounding. With regard to matching in case-control studies, it is clear that neither of the conditions of (4.2.4) can be met by matching. Since matching replaces the w_i's with the v_i's, (4.2.4b) reduces to

$$(b)' \quad \frac{(1-\varepsilon_i)}{(1-\delta_i)} = c \ , \quad \forall\, i = 1,\ldots,L \ .$$

If uniformity of the odds ratio is assumed, then (b)' reduces to

$$(b)'' \quad (1+\delta_i(1-OR))^{-1} = c \ , \quad \forall\, i = 1,\ldots,L$$

which cannot hold unless (a) holds. Only if the δ_i's and OR are "small" will (b)'' be true, and then only in an approximate sense.

4.2.3 Two Confounding Variables

It is useful to put the discussion in the preceding section into the context of two dichotomous confounding variables, F and G. The notation is altered by adding the subscript j to the parameters defined in Chapter 1, for example, $\alpha_{ij} = P(D/EF_iG_j)$ and $\theta_{ij} = P(F_iG_j/E)$. Expression (4.3.2) is now written as:

$$(a) \quad \beta_{ij} = \beta \quad \forall(i,j) \ ; \quad (b) \quad \theta_{ij} = \phi_{ij} \quad \forall(i,j) \ . \qquad (4.2.5)$$

Similarly, for the case-control study (4.3.3) is written as:

$$(a) \quad \delta_{ij} = \delta \quad \forall(i,j) \ ; \quad (b) \quad \frac{(1-\varepsilon_{ij})v_{ij}}{(1-\delta_{ij})w_{ij}} = c \quad \forall(i,j) \ . \qquad (4.2.6)$$

Expressions (4.3.5) and (4.3.6) can be rewritten in terms of odds ratios as follows:

$$\text{(a)}\quad \beta_{ij} = \beta,\ \forall(i,j) \Rightarrow \begin{cases} OR_{DF/\bar{E}G_j} = \dfrac{\beta_{1j}(1-\beta_{0j})}{\beta_{0j}(1-\beta_{ij})} = 1,\ j = 0,1 \\ OR_{DG/\bar{E}F_i} = \dfrac{\beta_{i1}(1-\beta_{i0})}{\beta_{i0}(1-\beta_{i1})} = 1,\ i = 0,1 \end{cases} \tag{4.2.7}$$

$$\text{(b)}\quad \theta_{ij} = \phi_{ij},\ \forall(i,j) \Rightarrow \begin{cases} OR_{EF/G_j} = \dfrac{\theta_{1j}\phi_{0j}}{\phi_{1j}\theta_{0j}} = 1,\ j = 0,1 \\ OR_{EG/F_i} = \dfrac{\theta_{i1}\phi_{i0}}{\phi_{i1}\theta_{i0}} = 1,\ i = 0,1 \end{cases} \tag{4.2.8}$$

$$\text{(c)}\quad \delta_{ij} = \delta,\ \forall(i,j) \Rightarrow \begin{cases} OR_{EF/\bar{D}G_j} = \dfrac{\delta_{1j}(1-\delta_{0j})}{\delta_{0j}(1-\delta_{1j})} = 1,\ j = 0,1 \\ OR_{EG/\bar{D}F_k} = \dfrac{\delta_{i1}(1-\delta_{i0})}{\delta_{i0}(1-\delta_{i1})} = 1,\ j = 0,1 \end{cases} \tag{4.2.9}$$

and

$$\text{(d)}\quad \frac{(1-\varepsilon_{ij})v_{ij}}{(1-\delta_{ij})w_{ij}} = c\ ,\ \forall(i,j) \Rightarrow OR_{DF/\bar{E}G_j} = OR_{DG/\bar{E}F_i} = 1,\ \forall(i,j)$$

where

$$\begin{cases} OR_{DF/\bar{E}G_j} = \dfrac{(1-\varepsilon_{1j})v_{1j}(1-\delta_{0j})w_{0j}}{(1-\delta_{1j})w_{1j}(1-\varepsilon_{0j})v_{0j}},\ j = 0,1\ . \\ OR_{DG/\bar{E}F_i} = \dfrac{(1-\varepsilon_{i1})v_{i1}(1-\delta_{i0})w_{i0}}{(1-\delta_{i1})w_{i1}(1-\varepsilon_{i0})v_{i0}},\ i = 0,1 \end{cases} \tag{4.2.10}$$

Sufficient conditions for no (joint) confounding in the case of two confounders are now apparent for both types of studies. In a follow-up study there is no confounding if disease is unrelated to

both F and G conditional on non-exposure and G_j, in the case of F, and F_i, in the case of G. Alternatively, if exposure is unrelated to both F and G conditional on G_j and F_i, respectively, then there is no joint confounding. Each of these conditions must obtain at both levels of the conditioning variables, F and G. For the case-control study the same relationships must hold except that the exposure association must be conditional further or non-disease. Table 4.1 summarizes these results.

Note that no conclusions regarding confounding are made for the case where one of F and G is (conditionally) unrelated to disease and the other is unrelated to exposure but neither F nor G is unrelated to both exposure and disease. In Section 4.3.5 this case is further evaluated by a regression formulation of no confounding.

The conditioning of the associations on particular levels of the confounding variables is extremely important. The relationships (or lack of) between disease and F, for example, and exposure and F must be conditioned on particular levels of G, the other confounder, as well as $\bar{E}$ or $\bar{D}$, as the case may be. With the introduction of a second extraneous variable G into the design, all of the previous conditions for no confounding must now obtain at each of the particular levels of G. This principle applies to both the exposure/confounder and disease/confounder relationships, outlined by (4.2.7 - 4.2.10). Hence, the former conditions for no confounding which, in effect, measure relationships after pooling over the second extraneous variable will not suffice and indeed may mislead concerning the actual nature of confounding in the data.

TABLE 4.1

Sufficient Conditions for No Confounding (Joint and Residual):
Two Extraneous Variables, F and G

	F and D unrelated[1]	F and E unrelated[2]
G and D unrelated[1]	no joint confounding	?
G and E unrelated[2]	?	no joint confounding

No residual confounding due to G

No residual confounding due to F

1. Conditional on non-exposure and F (in the case of G) or G (in the case of F).
2. For the follow-up study conditional on F (in the case of G) or G (in the case of F).
   For the case-control study conditional on non-disease and F (in the case of G) or G (in the case of F).

An example regarding the misleading conclusions which may be reached when considering only the marginal associations is given below. Let the two tables in Example 1 represent the joint distributions of F and G conditional on E and $\overline{E}$ in a follow-up study.

Example 1

θ_{ij}

	F_1	F_0	
G_1	.05	.35	.4
G_0	.25	.35	.6
	.3	.7	

ϕ_{ij}

	F_1	F_0	
G_1	.25	.15	.4
G_0	.05	.55	.6
	.3	.7	

In this case both marginal distributions agree but the joint distributions clearly do not, and there is potential confounding. In fact, if the stratum-specific RR's and β's are defined as below, then there over a 20% bias due to confounding.

β_{ij}

	F_1	F_0
G_1	0.01	0.002
G_0	0.005	0.001

Let $RR_{ij} = 2.5$, $\forall(i,j)$

$\Rightarrow$ sRR = 2.5

$$cRR = \frac{\left(\sum\sum_{ij} \beta_{ij}\theta_{ij}\right)}{\left(\sum\sum_{ij} \beta_{ij}\phi_{ij}\right)} \cdot sRR = \frac{0.00286}{0.00360}(2.5) = 1.986$$

$$\% \text{ bias} = \frac{sRR - cRR}{sRR} \times 100 = \frac{0.514}{2.5} = 20.6\%$$

The principle of evaluating the confounding property of an extraneous variable by conditioning on other known or suspected confounders has not been suggested in the literature, except by Fisher and Patil

(1974), who, nevertheless, do not fully identify the extent of conditioning necessary. Miettinen (1974), in a companion paper to Fisher and Patil's, argues that the marginal associations are sufficient to evaluate confounding and that rarely will the conditional associations reveal confounding when the marginal associations do not. Whether or not this is true in practice, Example 1 indicates the danger of following this reasoning.

Fisher and Patil's work advocates that "relatedness" between F and exposure in a case-control study be evaluated conditional on G but not on non-disease. In that regard an implicit "rare disease" assumption must be made in order that confounding be properly evaluated. Also, the exposure F relationship in a case-control study is presumably measured by pooling over cases and controls. In fact, this association cannot be properly measured unless the sampling ratio ($\rho = n_1/n_0$) and the prevalence, P(D), are known and taken into account. Even if known, unless the disease is rare, a serious error can be made by evaluating this association. The example which follows is an illustration of such an error.

In a case-control study the association between E and F conditional on G_j, $j = 0,1$ can be measured by the odds ratio, OR_{EF/G_j} defined below.

$$OR_{EF/G_j} = \frac{P(EF_1/G_j) \cdot P(\bar{E}F_0/G_j)}{P(EF_0/G_j) \cdot P(\bar{E}F_1/G_j)} = \frac{P(EF_1G_j) \cdot P(\bar{E}F_0G_j)}{P(EF_0G_j) \cdot P(\bar{E}F_1G_j)}$$

$$= \frac{[\varepsilon_{1j}v_{1j}\gamma+\delta_{1j}w_{1j}(1-\gamma)][(1-\varepsilon_{0j})v_{0j}\gamma+(1-\delta_{0j})w_{0j}(1-\gamma)]}{[(1-\varepsilon_{1j})v_{1j}\gamma+(1-\delta_{1j})w_{1j}(1-\gamma)][\varepsilon_{0j}v_{0j}\gamma+\delta_{0j}w_{0j}(1-\gamma)]} \quad (4.3.11)$$

where $\gamma = P(D)$.

Consider the following example where there is no confounding (and no effect modification).

Example 2

Stratum	v_{ij}	w_{ij}	ε_{ij}	δ_{ij}	OR_{ij}
(1,1)	0.40	0.05			
(1,0)	0.45	0.10	0.40	0.16	3.5
(0,1)	0.05	0.50			
(0,0)	0.10	0.45			

Also let $\gamma = 0.20$

Substituting into (4.2.11)

$$OR_{EF/G_1} = 2.34 \quad \text{and} \quad OR_{EF/G_0} = 1.93$$

But if conditioning on $\overline{D}$ as well, then

$$OR_{EF/\overline{D}G_1} = \frac{\delta_{11}(1-\delta_{01})}{\delta_{01}(1-\delta_{11})} = 1 \ ,$$

and

$$OR_{EF/\overline{D}G_0} = \frac{\delta_{10}(1-\delta_{00})}{\delta_{00}(1-\delta_{10})} = 1 \ .$$

On the basis of (4.2.11) one might have concluded wrongly that F was a potential confounder and considered adjustment for F in the analysis. (Indeed, F would have been classified a confounder since $OR_{DF/\overline{E}G_1} = OR_{DF/\overline{E}G_0} = 64$). Other examples can be constructed in which misleading notions regarding confounding are given by (4.2.11).

Examination of this criterion more closely under the assumption that the disease in question is rare reveals that (4.2.11) is not an unreasonable measure for the exposure/confounder association. Note that as $\gamma \to 0$, then

$$OR_{EF/G_j} \to OR_{EF/\bar{D}G_j} ,$$

and

$$OR_{EG/F_i} \to OR_{EG/\bar{D}F_i} .$$

Fisher and Patil are correct in suggesting that "relatedness" between a variable and exposure or disease should be evaluated conditional on other known or suspected confounders. Unfortunately, they choose to measure OR_{EF/G_j} by pooling cell frequencies over D and $\bar{D}$. By doing so they not only fail to take into consideration the prevalence but also the sampling ratio. Only if the sample sizes are equal and the prevalence is small will their suggested criterion for confounding be valid. Rather than resurrecting a correct assessment of OR_{EF/G_j}, etc., the expressions developed in this section should be adopted instead.

4.2.4 No Interaction Between the Confounding Variables

In a subsequent section of this chapter and in the next chapter, it will be of interest to consider the implications of imposing the restrictions of no interaction between F and G on exposure and disease. In this section conditions for no interaction between F and G are described in terms of equalities between odds ratios. The particular odds ratios representing associations between the

extraneous variables and exposure (or disease) are chosen primarily for their convenient application in later analyses.

In a follow-up study no interaction between F and G on exposure is expressed as $OR_{FG/E} = OR_{FG/\bar{E}}$, or

$$\frac{\theta_{11}\theta_{00}}{\theta_{01}\theta_{10}} = \frac{\phi_{11}\phi_{00}}{\phi_{10}\phi_{01}}, \tag{4.2.12}$$

which can also be expressed as $OR_{EF/G_1} = OR_{EF/G_0}$, or

$$\frac{\theta_{11}\phi_{01}}{\theta_{01}\phi_{11}} = \frac{\theta_{10}\phi_{00}}{\theta_{00}\phi_{10}}, \tag{4.2.13}$$

and as $OR_{EG/F_1} = OR_{EG/F_0}$, or

$$\frac{\theta_{11}\phi_{10}}{\theta_{10}\phi_{11}} = \frac{\theta_{01}\phi_{00}}{\theta_{00}\phi_{01}}. \tag{4.2.14}$$

With regard to disease no interaction between F and G can be expressed as $OR_{DF/\bar{E}G_1} = OR_{DF/\bar{E}G_0}$, or

$$\frac{\beta_{11}(1-\beta_{01})}{\beta_{01}(1-\beta_{11})} = \frac{\beta_{10}(1-\beta_{00})}{\beta_{00}(1-\beta_{10})}, \tag{4.2.15}$$

which can also be expressed as $OR_{DG/\bar{E}F_1} = OR_{DG/\bar{E}F_0}$, or

$$\frac{\beta_{11}(1-\beta_{10})}{\beta_{10}(1-\beta_{11})} = \frac{\beta_{01}(1-\beta_{00})}{\beta_{00}(1-\beta_{01})}. \tag{4.2.16}$$

In a case-control study no interaction between F and G on exposure implies that $OR_{EF/\bar{D}G_1} = OR_{EF/\bar{D}G_0}$, or

$$\frac{\delta_{11}(1-\delta_{01})}{\delta_{01}(1-\delta_{11})} = \frac{\delta_{10}(1-\delta_{00})}{\delta_{00}(1-\delta_{10})}, \tag{4.2.17}$$

or, equivalently that $OR_{EG/\bar{D}F_1} = OR_{EG/\bar{D}F_0}$, or

$$\frac{\delta_{11}(1-\delta_{10})}{\delta_{10}(1-\delta_{11})} = \frac{\delta_{01}(1-\delta_{00})}{\delta_{00}(1-\delta_{01})} . \qquad (4.2.18)$$

No interaction between F and G on disease is represented by

$OR_{DF/\bar{E}G_1} = OR_{DF/\bar{E}G_0}$

$$\frac{(1-\varepsilon_{11})v_{11}(1-\delta_{01})w_{01}}{(1-\delta_{11})w_{11}(1-\varepsilon_{01})v_{01}} = \frac{(1-\varepsilon_{10})v_{10}(1-\delta_{00})w_{00}}{(1-\delta_{10})w_{10}(1-\varepsilon_{00})v_{00}} \qquad (4.2.19)$$

or, equivalently by $OR_{DG/\bar{E}F_1} = OR_{DG/\bar{E}F_0}$

$$\frac{(1-\varepsilon_{11})v_{11}(1-\delta_{10})w_{10}}{(1-\delta_{11})w_{11}(1-\varepsilon_{10})v_{10}} = \frac{(1-\varepsilon_{01})v_{01}(1-\delta_{00})w_{00}}{(1-\varepsilon_{01})w_{01}(1-\delta_{00})v_{00}} . \qquad (4.2.20)$$

4.2.5 Regression Formulation of No Confounding

The conditions for no joint confounding can also be described by a multiple regression model for the risk of disease. Miettinen and Cook (1980) considered a model which expressed the risk of disease as a linear function of exposure and a single extraneous (potentially confounding) variable. In this section a similar model involving two potential confounding variables is developed. This model is not presented as the preferred model for describing the relationship between the risk of disease and exposure while incorporating other risk factors, but rather as a useful tool in which to clarify the concept of confounding as it involves multiple (potentially) confounding variables. The discussion is again limited in extent to dichotomous disease (D), exposure (E), and extraneous variables (F and G).

Let E, F, and G be indicator variables such that

$$E_k = \begin{cases} 1 & \text{if } k = 1 \quad \text{(exposure)} \\ 0 & \text{if } k = 0 \quad \text{(non-exposure)} \end{cases}$$

$$F_i = \begin{cases} 1 & \text{if } i = 1 \\ 0 & \text{if } i = 0 \end{cases} \qquad G_j = \begin{cases} 1 & \text{if } j = 1 \\ 0 & \text{if } j = 0 \end{cases}.$$

The conditional probability of exposure, $P(E_1/F_iG_j)$ can be written as

$$P(E_1/F_iG_j) = a + bF_1 + cG_j \tag{4.2.21}$$

where a, b, and c are regression parameters, and $P(E_1/F_iG_j)$ can be considered the expected value of E conditional on F_i and G_j.

In addition, the expected value of D conditional on F_i, G_j and E_k, $P(D/E_iG_jE_k)$ is formulated in terms of the following regression model

$$P(D/F_iG_jE_k) = \gamma_0 + \gamma_1F_i + \gamma_2G_j + \beta E_k \tag{4.2.22}$$

where γ_0, γ_1, γ_2, and β are regression parameters.

Both (4.2.21) and (4.2.22) assume that there is no interaction between F and G with respect to exposure and disease (Section 4.3.4).

Attention is focused on β in (4.2.22), the regression parameter which relates the risk of disease to exposure, in terms of the risk difference. Note that

$$\begin{aligned} P(D/E_1F_iG_j) - P(D/E_0F_iG_j) &= \gamma_0 + \gamma_1F_i + \gamma_2G_j + \beta \\ &\quad - (\gamma_0 + \gamma_1F_i + \gamma_2G_j) \\ &= \beta, \quad \forall(i,j) . \end{aligned}$$

Hence, β represents the uniform stratum-specific risk difference, RD_{ij}. There is no confounding if the expected cRD is equivalent to β. The cRD is now examined with respect to parameters of (4.2.21) and (4.2.22). First, let

$\pi_{ij} = P(F_i G_j)$, the unconditional joint distribution of F and G .

Then

$$cRD = P(D/E_1) - P(D/E_0)$$

$$= \frac{P(DE_1)}{P(E_1)} - \frac{P(DE_0)}{P(E_0)}$$

$$= [P(E_1)\cdot P(E_0)]^{-1}\{P(DE_1)\cdot P(E_0) - P(DE_0)\cdot P(E_1)\} . \qquad (4.2.23)$$

Now,

$$P(E_1) = \sum\sum_{ij} P(E_1/F_i G_j)\cdot P(F_i G_j)$$

$$= (a+b+c)\pi_{11} + (a+b)\pi_{10} + (a+c)\pi_{01} + a\pi_{00}$$

$$= a + b(\pi_{11} + \pi_{10}) + c(\pi_{11} + \pi_{01})$$

$$= \psi$$

$$P(DE_1) = \sum\sum_{ij} P(DE_1 F_i G_j) = \sum\sum_{ij} P(D/E_1 F_i G_j)\cdot P(E_1/F_i G_j)\cdot P(F_i G_j)$$

$$= (\gamma_0 + \gamma_1 + \gamma_2 + \beta)(a+b+c)\pi_{11} + (\gamma_0 + \gamma_1 + \beta)(a+b)\pi_{10}$$

$$+ (\gamma_0 + \gamma_2 + \beta)(a+c)\pi_{01} + (\gamma_0 + \beta)a\pi_{00}$$

and,

$$P(ED_0) = \sum_{ij}\sum P(DE_0F_iG_j) = \sum_{ij}\sum P(DE_0F_iG_j)\cdot P(E_0/F_iG_j)\cdot P(F_iG_j)$$

$$= (\gamma_0+\gamma_1+\gamma_2)(1-a-b-c)\pi_{11} + (\gamma_0+\gamma_1)(1-a-b)\pi_{10}$$

$$+ (\gamma_0+\gamma_2)(1-a-c)\pi_{01} + \gamma_0(1-a)\pi_{00} .$$

The term in brackets in (4.2.23) is now

$$[(\gamma_0+\gamma_1+\gamma_2+\beta)(a+b+c)\pi_{11}+(\gamma_0+\gamma_1+\beta)(a+b)\pi_{10}+(\gamma_0+\gamma_2+\beta)(a+c)\pi_{01}$$

$$+(\gamma_0+\beta)a\cdot\pi_{00}](1-\psi)$$

$$-[(\gamma_0+\gamma_1+\gamma_2)(1-a-b-c)\pi_{11}+(\gamma_0+\gamma_1)(1-a-b)\pi_{10}+(\gamma_0+\gamma_2)(1-a-c)\pi_{01}$$

$$+\gamma_0(1-a)\pi_{00}]\psi$$

$$= [(\gamma_0+\beta)(a+b(\pi_{11}+\pi_{10})+c(\pi_{11}+\pi_{01}))+\gamma_1(a+b)(\pi_{11}+\pi_{10})$$

$$+\gamma_2(a+c)(\pi_{11}+\pi_{01})+\pi_{11}(\gamma_1c+\gamma_2b)](1-\psi)$$

$$-[-\gamma_0(a+b(\pi_{11}+\pi_{10})+c(\pi_{11}+\pi_{01}))-\gamma_1(a+b)(\pi_{11}+\pi_{10})-\gamma_0(a+c)(\pi_{11}+\pi_{01})$$

$$+\gamma_0+\gamma_1(\pi_{11}+\pi_{10})+\gamma_2(\pi_{11}+\pi_{01})-\pi_{11}(\gamma_1c+\gamma_2b)]\psi$$

$$= [(\gamma_0+\beta)\psi+\gamma_1(a+b)(\pi_{11}+\pi_{10})+\gamma_2(a+c)(\pi_{11}+\pi_{01})$$

$$+(\gamma_1c+\gamma_2b)\pi_{11}](1-\psi)$$

$$-[\gamma_0(1-\psi)+\gamma_1(1-a-b)(\pi_{11}+\pi_{10})+\gamma_2(1-a-c)(\pi_{11}+\pi_{01})$$

$$-(\gamma_1c+\gamma_2b)\pi_{11}]\psi$$

$$= \beta\psi(1-\psi) + \gamma_1((a+b)(1-\psi) - \psi(1-a-b))(\pi_{11}+\pi_{10})$$

$$+ \gamma_2((a+c)(1-\psi) - \psi(1-a-c))(\pi_{11}+\pi_{01})$$

$$+ (\gamma_1 c + \gamma_2 b)(\pi_{11}(1-\psi) - \pi_{11}\psi)$$

$$= \beta\psi(1-\psi) + \gamma_1(a+b-\psi)(\pi_{11}+\pi_{10}) + \gamma_2(a+c-\psi)(\pi_{11}+\pi_{01})$$

$$+ (\gamma_1 c + \gamma_2 b)\pi_{11} \ .$$

And, since

$$a+b-\psi = a+b - (a+b(\pi_{11}+\pi_{10}) + c(\pi_{11}+\pi_{01}))$$

$$= b(\pi_{01}+\pi_{00}) - c(\pi_{11}+\pi_{01})$$

and,

$$a+c-\psi = a+c - (a+b(\pi_{11}+\pi_{10}) + c(\pi_{11}+\pi_{01}))$$

$$= c(\pi_{10}+\pi_{00}) - b(\pi_{11}+\pi_{10}) \ ,$$

$$P(DE_1)\cdot P(E_0) - P(DE_0)\cdot P(E_1) = \beta\psi(1-\psi) + \gamma_1(b(\pi_{01}+\pi_{00}) - c(\pi_{11}+\pi_{01}))(\pi_{11}+\pi_{10})$$

$$+ \gamma_2(c(\pi_{10}+\pi_{00}) - b(\pi_{11}+\pi_{10}))(\pi_{11}+\pi_{01})$$

$$+ \pi_{11}(\gamma_1 c + \gamma_2 b)$$

$$= \beta\psi(1-\psi) + \gamma_1 b(\pi_{11}+\pi_{10})(\pi_{01}+\pi_{00})$$

$$+ \gamma_2 c(\pi_{11}+\pi_{01})(\pi_{10}+\pi_{00})$$

$$+ (\gamma_1 c + \gamma_2 b)\{\pi_{11} - (\pi_{11}+\pi_{01})(\pi_{11}+\pi_{10})\} \quad (4.2.24)$$

The expression in brackets reduces to

$$\pi_{11}(1-\pi_{11}-\pi_{01}-\pi_{10}) - \pi_{01}\pi_{10} = \pi_{11}\pi_{00} - \pi_{01}\pi_{10} \ , \quad (4.2.25)$$

the cross-products difference, which can also be written as

$$\pi_{01}\pi_{10}(OR_{FG}-1) \quad , \quad \text{where} \quad OF_{FG} = \frac{\pi_{11}\pi_{00}}{\pi_{10}\pi_{01}} \; .$$

Therefore, dividing (4.2.24) by $\psi(1-\psi)$ and substituting (4.2.25), the expression (4.2.23) for the cRD becomes

$$\begin{aligned} cRD &= \beta + \frac{\gamma_i b(\pi_{11}+\pi_{10})(1-\pi_{11}-\pi_{10})}{\psi(1-\psi)} + \frac{\gamma_2 c(\pi_{11}+\pi_{01})(1-\pi_{11}-\pi_{01})}{\psi(1-\psi)} \\ &\quad + \frac{(\gamma_1 c+\gamma_2 b)\pi_{01}\pi_{10}(OR_{FG}-1)}{\psi(1-\psi)} \\ &= \beta + f(\gamma_1,\gamma_2,b,c,OR_{FG},\underset{\sim}{\pi},\psi) \; . \end{aligned} \tag{4.2.26}$$

The regression parameters, b and c, represent measures of conditional association between F and E, and G and E, respectively. The condition: $b = 0$, can be interpreted as "no association between F and E conditional on G," and likewise for the condition: $c = 0$. In addition, γ_1 and γ_2 represent measures of conditional association between F and D, and G and D, respectively. If $\gamma_1 = 0$, then there is "no association between F and D conditional on E and G," and likewise for $\gamma_2 = 0$.

The function, $f(\cdot)$, reveals the complex nature of joint confounding when more than one extraneous variable is involved. There is no confounding if $f(\cdot) = 0$. No necessary <u>and</u> sufficient conditions for no confounding are implied by $f(\cdot) = 0$. Indeed, it is apparent that the presence or absence of confounding is dependent not only on the relationships of each extraneous variable to exposure and disease but also on their intercorrelation.

If one can assume that F and G are independent ($OR_{FG} = 1$), then there is no confounding if one of the following conditions is met:

(a) $b = 0$ and $c = 0 \Rightarrow OR_{EF/G} = OR_{EG/F} = 1$ (which agrees with (4.2.8)),

(b) $\gamma_1 = 0$ and $\gamma_2 = 0 \Rightarrow OR_{DF/EG} = OR_{DG/EF} = 1$ (which agrees with (4.2.7)), or

(c) either $\gamma_1 = 0$ and $c = 0$ or $\gamma_2 = 0$ and $b = 0$,
$\Rightarrow OR_{DF/EG} = 1$ and $OR_{EG/F} = 1$, or $OR_{DG/EF} = 1$ and $OR_{EF/G} = 1$

That is, if F and G are independent, then there is no confounding if each variable is unrelated to either exposure or disease. This statement of no confounding resembles the prevailing description of a non-confounder. However, that description breaks down when a second confounder which is correlated to the first is introduced. In that circumstance there may still be confounding even though each variable meets the individual requirements for no confounding.

As a result not only must the relationships between a potential confounder and exposure and disease be evaluated but also relationships between that variable and other known confounders must be evaluated, in order to properly characterize the confounding potential of a variable.

Expression (4.2.26) provides additional insight into the nature of confounding. The direction of confounding is determined by the "direction" of the confounder as well as the direction of the F-G association. A confounder can be called "positive" if it is directly (or, inversely) associated with both D and E. Otherwise, it can be called a "negative" confounder. If $\gamma_1 > 0$ and $b > 0$, which implies

$\gamma_1 b > 0$, then F is a "positive" confounder, and the confounding associated with F is positive. If $\gamma_1 > 0$ and $b < 0$ or $\gamma_1 < 0$ and $b > 0$, then $\gamma_1 b < 0$ and the confounding associated with F is negative. Of course, the other factor G must be considered; they cannot be treated in isolation unless they are independent.

Another factor which affects the strength of the confounding is the marginal distributions of F and G (represented by $(\pi_{11} + \pi_{10})$ and $(\pi_{11} + \pi_{01})$, respectively). When the marginal distribution of F is skewed, the strength of the confounding due to that variable is diminished. Confounding is maximized with respect to each marginal distribution as $p \to 0.5$, where $p = \pi_{11} + \pi_{10}$ (or $\pi_{11} + \pi_{01}$), or as the distribution becomes more uniform. An implication of this result is that control for a variable which otherwise meets all of the requirements for a confounder may not be necessary if that variable is highly skewed in distribution.

4.3 Residual Confounding

The discussion leading up to this section has concerned the concepts of "joint" and "marginal" confounding. The addition of a second extraneous variable to the scenario requires that the conditions for no confounding (marginally) hold at each level of each of the other confounders for there to be no (joint) confounding. However, even these conditions may not be sufficient for no confounding. In certain circumstances the extraneous variables may also have to be independent for there to be no confounding.

The developments of this chapter do not lead to a completely satisfactory description of confounding on a variable-by-variable basis. Only if the confounding effects of all other extraneous variables are removed, can the confounding potential of a particular variable be studied. In this regard a question which may arise is, as follows: if adjustment for the confounding effects of F is made, what are the conditions for no additional confounding due to G? In other words, what are the conditions for no "residual" confounding?

4.3.1 Follow-up Study

Let RR_F be the risk ratio adjusted for F marginally and $RR_{F,G}$ be the risk ratio adjusted for F and G. There is no residual confounding due to G if $RR_F = RR_{F,G}$. If the adjustments are internal standardizations, then $RR_{F,G}$ equals sRR and RR_F is given by:

$$RR_F = \left(\sum_i \alpha_i \theta_i\right) / \left(\sum_i \beta_i \theta_i\right)$$

where

$$\alpha_i = \frac{\alpha_{i1}\theta_{i1} + \alpha_{i0}\theta_{i0}}{\theta_{i1} + \theta_{i0}} = P(D/EF_i)$$

$$\beta_i = \frac{\beta_{i1}\phi_{i1} + \beta_{i0}\phi_{i0}}{\phi_{i1} + \phi_{i0}} = P(D/\bar{E}F_i) \ ,$$

and

$$\theta_i = \theta_{i1} + \theta_{i0} \qquad = P(F_i/E) \ .$$

$$RR_{F_1G} = RR_F \Rightarrow \frac{\sum_{ij}\sum \alpha_{ij}\theta_{ij}}{\sum_{ii}\sum \beta_{ij}\theta_{ij}} = \frac{\sum_i \left[\frac{\alpha_{i1}\theta_{i1} + \alpha_{i0}\theta_{i0}}{\theta_{i1} + \alpha_{i0}}\right](\theta_{i1} + \theta_{i0})}{\sum_i \left[\frac{\beta_{i1}\phi_{i1} + \beta_{i0}\phi_{i0}}{\phi_{i1} + \phi_{i0}}\right](\theta_{i1} + \theta_{i0})} \ ,$$

or

$$\sum_{ij}\sum \beta_{ij}\theta_{ij} = \sum_{i}\left[\frac{\beta_{i1}\phi_{i1}+\beta_{i0}\phi_{10}}{\phi_{i1}+\phi_{i0}}\right](\theta_{i1}+\theta_{i0}) = \sum_{ij}\sum \beta_{ij}\phi_{ij}\,\frac{(\theta_{i1}+\theta_{i0})}{(\phi_{i1}+\phi_{i0})}$$

which can be expressed as,

$$\sum_{ij}\sum \beta_{ij}\left[\theta_{ij} - \phi_{ij}\,\frac{(\theta_{i1}+\theta_{i0})}{(\phi_{i1}+\phi_{i0})}\right] = 0 ,$$

or

$$\sum_{i}\beta_{i1}\left[\frac{\theta_{i1}\phi_{i0}-\phi_{i1}\theta_{i0}}{\phi_{i1}+\phi_{i0}}\right] - \sum_{i}\beta_{i0}\left[\frac{\theta_{i1}\phi_{i0}-\phi_{i1}\theta_{i0}}{\phi_{i1}+\phi_{i0}}\right] = 0 ,$$

or

$$\sum_{i}(\beta_{i1}-\beta_{i0})\left[\frac{\theta_{i1}\phi_{i0}-\theta_{i0}\phi_{i1}}{\phi_{i1}+\phi_{i0}}\right] = 0 . \qquad (4.3.1)$$

An endless number of solutions to (4.3.1) can be derived; however, the following solutions are of special interest and interpretation. Expression (4.3.1) will hold if:

(a) $\beta_{i1} = \beta_{i0}$, $i = 0,1 \Rightarrow OR_{DG/\bar{E}F_i} = 1$, $i = 0,1$

(b) $\theta_{i1}\phi_{i0} = \theta_{i0}\phi_{i1}$, $i = 0,1 \Rightarrow OR_{EG/F_i} = 1$, $i = 0,1$

(c) $\beta_{11} = \beta_{10}$ and $\theta_{01}\phi_{00} = \theta_{00}\phi_{01} \Rightarrow \begin{cases} OR_{DG/\bar{E}F_1} = 1 \\ OR_{EG/F_0} = 1 \end{cases}$

or

(d) $\beta_{01} = \beta_{00}$ and $\theta_{11}\phi_{10} = \theta_{10}\phi_{11} \Rightarrow \begin{cases} OR_{EG/\bar{E}F_0} = 1 \\ OR_{EG/F_1} = 1 \end{cases}$ (4.3.2)

Hence, there is no residual confounding due to G in a follow-up study if either:

(a) there is no association between G and disease conditional on $\bar{E}$ and F ($OR_{GD/\bar{E}F_i} = 1$, $i = 0,1$);

(b) there is no association between G and exposure conditional on F ($OR_{GE/F_i} = 1$, $i = 0,1$); or

(c) at each level of F (conditional on F) there is either no association between G and disease conditional on $\bar{E}$ or no association between G and exposure ($OR_{GD/\bar{E}F_i} = OR_{EG/F_k} = 1$, $i \neq k = 0,1$).

In other words there is no residual confounding due to a variable G if, conditional on any other confounder, G is either unrelated to disease (conditional on non-exposure) or unrelated to exposure. Furthermore, if G is unrelated to either exposure or disease at each level of F, then there is no residual confounding due to G. To be concise if G if independent of either disease or exposure at every level of F, then there will be no confounding due to G. The above conditions are sufficient for no confounding but not necessary.

From earlier results in this chapter it is clear that conditions for no residual confounding can but will not necessarily lead to conditions for no joint confounding. Indeed, sufficient conditions for no residual confounding due to F and also do to G may not lead to sufficient conditions for no joint confounding. The empty cells in Table 4.1 represent the unknown state of nature with regard to joint confounding when there is no residual confounding. The following example demonstrates this phenomenon.

Example 3

Let the parameters for a follow-up study β_{ij}, θ_{ij}, and ϕ_{ij} be defined as below. The parameter α_{ij} is defined by $\alpha_{ij} = RR \cdot \beta_{ij}$,

where RR is the uniform, stratum-specific risk ratio.

Stratum	β_{ij}	θ_{ij}	ϕ_{ij}
(1,1)	0.003	0.45	0.19
(1,0)	0.003	0.15	0.26
(0,1)	0.0005	0.10	0.04
(0,0)	0.0005	0.30	0.51

Now,

$$cRR = \frac{\left(\sum\sum_{ij}\beta_{ij}\theta_{ij}\right)}{\left(\sum\sum_{ij}\beta_{ij}\phi_{ij}\right)} \cdot sRR = \frac{0.002}{0.001625} \cdot sRR = (1.231)\ sRR\ ,$$

which implies a 23.1% upward bias due to confounding.

Note, however, that

$$OF_{DG/\bar{E}F_i} = 1\ ,\quad i = 0,1 \quad \text{and} \quad OR_{EF/G_j} = \begin{cases} 0.95, & j = 1 \\ 0.98, & j = 1 \end{cases}\ .$$

There is no residual confounding due to F and to G; nevertheless, there is (joint) confounding. (Recall that under this circumstance, F and G must be independent. In fact, they are not since $OR_{FG/E} = 9$ and $OR_{FG/\bar{E}} = 9.32$.)

Interestingly, confounding can be controlled by adjusting for either F or G. Kleinbaum and Kupper (1980) refer to this phenomenon in their discussion of a "sufficient-confounder-group," which they define to be a minimal set of potential confounders requiring adjustment in the analysis to control confounding. In this instance both F and G individually would meet the requirements of sufficient-confounder-groups.

If there is no interaction between F and G, then necessary and sufficient conditions exist for no residual confounding. Consider expression (4.3.1), which can be written as:

$$\sum_i \frac{\beta_{i0}\theta_{i0}\phi_{i1}}{\theta_{i1}+\theta_{i0}} \left(\frac{\beta_{i1}}{\beta_{i0}} - 1\right)\left(\frac{\theta_{i1}\phi_{i0}}{\theta_{i0}\phi_{i1}} - 1\right) = 0 \ ,$$

or as

$$\sum_i \frac{\beta_{i0}(1-\beta_{i0})\theta_{i0}\phi_{i1}}{(\theta_{i1}+\theta_{i0})(1-\beta_{i0}(1-OR_{DG/\bar{E}F_i}))} (OR_{DG/\bar{E}F_i} - 1)(OR_{EG/F_i} - 1) = 0 \ ,$$

since

$$\beta_{i1} = \frac{\beta_{i0} \cdot OR_{DG/\bar{E}F_i}}{(1-\beta_{i0}(1-OR_{DG/\bar{E}F_i}))} \ .$$

Therefore, by the assumption of no interaction,

$$(OR_{DG/\bar{E}F}-1)(OR_{EG/F}-1) \sum_i \frac{\beta_{i0}(1-\beta_{i0})\theta_{i0}\phi_{i1}}{(\theta_{i1}+\theta_{i0})(1-\beta_{i0}(1-OR_{DG/EF_i}))} = 0 \ .$$

Since the summation term is always positive, then either:

(a) $OR_{DG/\bar{E}F} = 1$

or

(b) $OR_{EG/F} = 1$

for there to be no residual confounding due to G. These conditions, which correspond to (4.3.2a) and (4.3.2b), are now necessary and sufficient.

The sufficiency of the conditions for no residual confounding does not change the nature of the relationship between no residual and no

joint confounding, however. In the previous example F and G do not interact on either exposure or disease, there is no residual confounding, yet there remains joint confounding.

4.3.2 Case-Control Study

Analogous to the follow-up study, the conditions for no residual confounding are derived by comparing the odds ratio adjusted for F to that adjusted for F and G. Assuming that direct standardization is the method of adjustment,

$$sOR_F = \left(\sum_i \varepsilon_i v_i\right) \Big/ \sum_i \frac{(1-\varepsilon_i) v_i \delta_i}{1-\delta_i} \quad , \quad \text{where} \begin{cases} \varepsilon_i = \dfrac{\varepsilon_{i1} v_{i1} + \varepsilon_{i0} v_{i0}}{v_{i1} + v_{i0}} \\ \delta_i = \dfrac{\delta_{i1} w_{i1} + \delta_{i0} w_{i0}}{w_{i1} + w_{i0}} \\ v_i = v_{i1} + v_{i0} \, . \end{cases}$$

$$\Rightarrow sOR_F = \frac{\sum_i \left[\dfrac{\varepsilon_{i1} v_{i1} + \varepsilon_{i0} v_{i0}}{v_{i1} + v_{i0}}\right] (v_{i1} + v_{i0})}{\sum_i \left\{ \dfrac{\left[1 - \dfrac{\varepsilon_{i1} v_{i1} + \varepsilon_{i0} v_{i0}}{v_{i1} + v_{i0}}\right] (v_{i1} + v_{i0}) \left[\dfrac{\delta_{i1} w_{i1} + \delta_{i1} w_{i0}}{w_{i1} + w_{i0}}\right]}{\left[1 - \dfrac{\delta_{i1} w_{i1} + \delta_{i0} w_{i0}}{w_{i1} + w_{i0}}\right]} \right\}}$$

and

$$sOR_{F,G} = \left(\sum_{ij} \varepsilon_{ij} v_{ij}\right) \Big/ \sum\sum \left[\frac{(1-\varepsilon_{ij}) v_{ij} \delta_{ij}}{1-\delta_{ij}}\right] .$$

There is no residual confounding if $sOR_F = SOR_{F,G}$. Since the numerators of sOR_F and $sOR_{F,G}$ are equal, then

$$\sum_i \left\{ \frac{[(1-\varepsilon_{i1})v_{i1}+(1-\varepsilon_{i0})v_{i0}][\delta_{i1}w_{i1}+\delta_{i0}w_{i0}]}{[(1-\delta_{i1})w_{i1}+(1-\delta_{i0})w_{i0}]} \right\} = \sum_{ij}\sum \left[\frac{(1-\varepsilon_{ij})v_{ij}\delta_{ij}}{1-\delta_{ij}} \right] .$$

$$\Rightarrow (1-\varepsilon_{11})v_{11}\left(\frac{\delta_{11}w_{11}+\delta_{10}w_{10}}{(1-\delta_{11})w_{11}+(1-\delta_{10})w_{10}} - \frac{\delta_{11}}{1-\delta_{11}} \right)$$

$$+ (1-\varepsilon_{10})v_{10}\left(\frac{\delta_{11}w_{11}+\delta_{10}w_{10}}{(1-\delta_{11})w_{11}+(1-\delta_{10})w_{10}} - \frac{\delta_{10}}{1-\delta_{10}} \right)$$

$$+ (1-\varepsilon_{01})v_{01}\left(\frac{\delta_{01}w_{01}+\delta_{00}w_{00}}{(1-\delta_{01})w_{01}+(1-\delta_{00})w_{00}} - \frac{\delta_{01}}{1-\delta_{01}} \right)$$

$$+ (1-\varepsilon_{00})v_{00}\left(\frac{\delta_{01}w_{01}+\delta_{00}w_{00}}{(1-\delta_{01})w_{01}+(1-\delta_{00})w_{00}} - \frac{\delta_{00}}{1-\delta_{00}} \right) = 0 . \quad (4.3.3)$$

The first term in (4.3.3) can be expressed as

$$\frac{(1-\varepsilon_{11})v_{11}}{1-\delta_{11}} \left[\frac{\delta_{11}w_{11}(1-\delta_{11})+\delta_{10}w_{10}(1-\delta_{11})-\delta_{11}w_{11}(1-\delta_{11})-\delta_{11}(1-\delta_{10})w_{10}}{(1-\delta_{11})w_{11}+(1-\delta_{10})w_{10}} \right]$$

$$= \frac{(1-\varepsilon_{11})v_{11}}{1-\delta_{11}} \quad \frac{(\delta_{10}-\delta_{11})w_{10}}{(1-\delta_{11})w_{11}+(1-\delta_{10})w_{10}}$$

The remaining three terms can be expressed in the same fashion. Therefore, (4.3.3) leads to

$$\frac{(1-\varepsilon_{11})v_{11}(\delta_{10}-\delta_{11})w_{10}}{(1-\delta_{11})[(1-\delta_{11})w_{11}+(1-\delta_{10})w_{10}]} + \frac{(1-\varepsilon_{10})v_{10}(\delta_{11}-\delta_{10})w_{11}}{(1-\delta_{10})[(1-\delta_{11})v_{11}+(1-\delta_{10})w_{10}]}$$

$$+ \frac{(1-\varepsilon_{01})v_{01}(\delta_{01}-\delta_{00})w_{00}}{(1-\delta_{01})[(1-\delta_{01})w_{01}+(1-\delta_{00})w_{00}]} + \frac{(1-\varepsilon_{00})v_{00}(\delta_{00}-\delta_{01})w_{01}}{(1-\delta_{00})[(1-\delta_{01})w_{01}+(1-\delta_{00})w_{00}]} = 0$$

or,

$$\sum_i \left[\frac{(1-\varepsilon_{i1})v_{i1}}{(1-\delta_{i1})w_{i1}} - \frac{(1-\varepsilon_{i0})v_{i0}}{(1-\delta_{i0})w_{i0}} \right] \frac{w_{i1}w_{i0}(\delta_{i1}-\delta_{i0})}{((1-\delta_{i1})w_{i1}+(1-\delta_{i0})w_{i0})} = 0 . \quad (4.3.4)$$

Four conditions of particular interest can be established from (4.3.4) and which imply no residual confounding:

(a) $\delta_{i1} = \delta_{i0}$, $i = 0,1 \Rightarrow OR_{EG/\bar{D}F_i} = 1$, $i = 0,1$

(b) $\dfrac{(1-\varepsilon_{i1})v_{i1}}{(1-\delta_{i1})w_{i1}} = \dfrac{(1-\varepsilon_{i0})v_{i0}}{(1-\delta_{i0})w_{i0}} \Rightarrow OR_{DG/\bar{E}F_i} = 1$, $i = 0,1$

(c) $\delta_{11} = \delta_{10}$ and $\dfrac{(1-\varepsilon_{00})v_{00}}{(1-\delta_{00})w_{00}} = \dfrac{(1-\varepsilon_{00})v_{00}}{(1-\delta_{00})w_{00}} \Rightarrow \begin{cases} OR_{EG/\bar{D}F_1} = 1 \\ OR_{DG/\bar{E}F_0} = 1 \end{cases}$

(d) $\delta_{01} = \delta_{00}$ and $\dfrac{(1-\varepsilon_{11})v_{11}}{(1-\delta_{11})v_{11}} = \dfrac{(1-\varepsilon_{10})v_{10}}{(1-\delta_{10})w_{10}} \Rightarrow \begin{cases} OR_{EG/\bar{D}F_0} = 1 \\ OR_{DG/\bar{E}F_1} = 1 \end{cases}$

These conditions are identical to those derived for the follow-up study with the exception that the association between exposure and G must be measured from the controls only. And, just as for the follow-up study, these conditions are sufficient only for no residual confounding. Should these conditions hold for both F and G, there still may be (joint) confounding.

Imposing the restrictions implied by no interaction between F and G will lead to necessary conditions for no residual confounding. From (4.3.4) we have

$$\sum_i \frac{\delta_{i0}(1-\varepsilon_{i0})v_{i0}}{(1-\delta_{i0})w_{i0}} \left[\frac{(1-\varepsilon_{i1})v_{i1}(1-\delta_{i0})v_{i0}}{(1-\delta_{i1})w_{i1}(1-\varepsilon_{i0})v_{i0}} - 1 \right] \frac{w_{i1}w_{i0}}{(1-\delta_{i1})w_{i1}+(1-\delta_{i0})w_{i0}} \left(\frac{\delta_{i1}}{\delta_{i0}} - 1 \right) = 0$$

Since

$$\delta_{i1} = OR_{EG/\bar{D}F_k} \cdot \delta_{i0}/(1-\delta_{i0}(1-OR_{EG/\bar{D}F_i}))$$

$$\Rightarrow \frac{\delta_{i1}}{\delta_{i0}} - 1 = (OR_{EG/\bar{D}F_i} - 1)(1 - \delta_{i0}) ,$$

then

$$\sum_i \frac{\delta_{i0}(1-\varepsilon_{i0})v_{i0}w_{i1}}{(1-\delta_{i0}(1-OR_{EG/\bar{D}F_i}))[(1-\delta_{i1})w_{i1}+(1-\delta_{i0})w_{i0}]} (OR_{EG/\bar{D}F_i}-1) \cdot (OR_{DG/\bar{E}F_i}) = 0 .$$

Therefore, by the assumption of no interaction,

$$(OR_{EG/\bar{D}F}-1)(OR_{DG/\bar{E}F}-1)\sum_i \frac{\delta_{i0}(1-\varepsilon_{i0})v_{i0}w_{i1}}{(1-\delta_{i0}(1-OR_{EG/\bar{D}F_i}))[(1-\delta_{i1})w_{i1}+(1-\delta_{i0})w_{i0}]} = 0 . \tag{4.3.5}$$

The summation term is always positive; hence, either

(a) $OR_{EG/\bar{D}F} = 1$, or

(b) $OR_{DG/\bar{E}F} = 1$ (4.3.6)

for (4.3.5) to hold. The necessary and sufficient conditions for no residual confounding in a case-control study are that there is no association between G and disease conditional on non-exposure and F or no association between G and exposure conditional on non-disease and F. If the disease can be considered rare, then conditioning on non-disease is equivalent to no conditioning on disease whatsoever, and the conditions for no residual confounding under these circumstances are the same as those for the follow-up study.

Just as for the follow-up study the sufficiency of (4.3.6) for no residual confounding will not guarantee no joint confounding. That is, if neither F nor G are residual confounders, and if, in addition, F and G do not interact on exposure and disease, there still may be (joint) confounding. The size and direction of the confounding will depend upon the nature of the association between F and G.

4.4 Validity and the Analysis of Matched Data

In epidemiology matching has long been considered the primary method for controlling confounding. Until recently, however, the

efficacy of matching with regard to the control of confounding had not been studied analytically. Kupper, et al. (1980) showed that matching could control confounding in follow-up studies but not in case-control studies; that is, stratification or some other method of adjustment would have to follow matching in the analysis in order to ensure the control of confounding.

The fact that matching in and of itself fails to control confounding does not result from the case-control design but rather from the measure of effect associated with the case-control design, the odds ratio. Should the odds ratio be chosen for the measure of effect in the follow-up study where matching is employed, then there would still remain confounding as well. The crude matched odds ratio (cORm) cannot be expressed in terms of a weighted average of stratum-specific odds ratios and, therefore, is not free of confounding bias. Nevertheless, the question arises: how strong is the confounding bias associated with this measure of effect? Is the cORm an improvement over the cOR in terms of reducing confounding? And, more generally, what is the appropriate method of analysis of matched data?

In this section the question is considered from two points of view. From the first, the matching variable is the only confounder, and the concern is whether analytical adjustment of the data is needed in a practical sense. From the second, the matching variable is not the only confounder; hence, which variables the analytical adjustment must involve, only the unmatched confounder or both confounders, is studied.

4.4.1 One Matching Variable and Confounder

In order to examine whether matching in and of itself controls confounding in a practical sense, cORm was evaluated numerically across a range of values, assuming uniformity of the stratum-specific odds ratios. The distribution of the confounding variable conditional on disease was allowed to range from (0.9, 0.1) to (0.1, 0.9) in increments of 0.1. Exposure probabilities among the controls were allowed to range from 0.01 to 0.36 and the odds ratio from 1.5 to 5. In addition, $OR_{EF/\bar{D}}$ was allowed to range from 2 to 5.

The underlying value of the parameter $OR_{DF/\bar{E}}$ has no real consequence on the value of cORm since the matching process forces the parameter to assume a value slightly less than 1. Matching replaces the w_i by the v_i in the analysis; hence, after matching

$$OR_{DF/\bar{E}} = \frac{(1-\epsilon_1)(1-\delta_0)}{(1-\delta_1)(1-\epsilon_0)} = \frac{1+\delta_0(OR-1)}{1+\delta_1(OR-1)}$$

$$\leq 1 \ , \quad \text{since} \quad OR_{EF/\bar{D}} \geq 1 \quad \text{by assumption} \ .$$

The value of $OR_{DF/\bar{E}}$ is now only a reflection of $OR_{EF/\bar{D}}$ and the odds ratio. The underlying value of $OR_{DF/\bar{E}}$ has direct bearing on the magnitude of cOR, however. In Tables 4.2 and 4.3 the range of possible cOR is given for each cORm. After fixing the other parameters, a particular value of cOR is determined by $\underset{\sim}{w} = (w_1, w_0)$, the distribution of F among the controls, which is not specified by a given cORm. The minimum value of cOR occurs when $w_1 = 1$, corresponding to $OR_{DF/\bar{E}} = 0$, and the maximum occurs when $w_0 = 1$, corresponding to an infinite $OR_{DF/\bar{E}}$.

It can be shown analytically that cORm is always biased toward the null. (See Appendix 3.) The results of the numerical evaluations indicate that this bias is generally not severe. For small to moderate odds ratios, cORm is almost free of confounding bias except when the exposure probabilities are large; when the exposure probabilities are small ($\delta_i < 0.1$), the bias does not exceed 2%. The bias increases with increasing odds ratio, exposure probabilities, and $OR_{EF/\overline{D}}$. At large values of the odds ratio, high exposure probabilities, and large $OR_{EF/\overline{D}}$ (potentially large confounding bias), cORm can be quite misleading. For example, if $OR = 5$, $\delta_0 = 0.1$, and $OR_{EF/D} = 5$ (the exposure probabilities range up to 0.74), the bias can reach 19% (see Table 4.3). Unless these contingencies occur simultaneously, however, the bias will rarely exceed 6%.

In order the evaluate whether cORm effectively reduces confounding, it must be compared to cOR under random sampling. The results of a number of evaluations are summarized by Tables 4.2 and 4.3. Contrasting cORm to the range of possible cOR under a wide variety of values of the population parameters indicates that cORm is indeed effective in terms of reducing confounding. While cORm may be biased toward the null by over 10% in extreme cases, the corresponding cOR are biased over 200%. Other values of cOR can be biased beyond 300% of the true OR. Of course, within the range of values for cOR there is a value which equals the true (uniform) OR. For the circumstanes where cOR = OR (i.e., no confounding), then matching actually introduces confounding, since cORm < OR. However, in most situations this bias is not large enough to obscure an accurate assessment of the association between exposure and disease via a crude matched analysis.

TABLE 4.2

Comparison of the Crude Matched Odds Ratio (cORm) With the True, Uniform Odds Ratio (OR) and the Crude Odds Ratio (cOR): Small to Moderate OR

			$OR_{EF/\overline{D}} = 2$				$OR_{EF/\overline{D}} = 5$			
			$\delta_0 = 0.01$		$\delta_0 = 0.10$		$\delta_0 = 0.01$		$\delta_0 = 0.10$	
OR	v_1	v_0	cORm	cOR^1	cORm	cOR^1	cORm	cOR^1	cORm	cOR^1
1.5	.1	.9	1.500	(1.424,2.848)	1.497	(1.415,2.831)	1.498	(1.374, 6.868)	1.481	(1.322,6.608)
	.3	.7	1.499	(1.273,2.545)	1.493	(1.253,2.507)	1.494	(1.125, 5.626)	1.453	(1.017,5.085)
	.5	.5	1.499	(1.122,2.244)	1.490	(1.100,2.200)	1.490	(0.883, 4.413)	1.437	(0.767,3.833)
	.7	.3	1.499	(0.973,1.945)	1.491	(0.955,1.909)	1.489	(0.646, 3.228)	1.438	(0.557,2.786)
	.9	.1	1.499	(0.824,1.648)	1.496	(0.316,1.633)	1.493	(0.414, 2.069)	1.465	(0.379,1.896)
2.25	.1	.9	2.249	(2.135,4.271)	2.240	(2.118,4.235)	2.241	(2.056,10.279)	2.185	(1.950,9.750)
	.3	.7	2.247	(1.907,3.815)	2.224	(1.868,3.736)	2.226	(1.677, 8.386)	2.094	(1.466,7.331)
	.5	.5	2.245	(1.681,3.363)	2.217	(1.636,3.273)	2.215	(1.312, 6.559)	2.049	(1.093,5.461)
	.7	.3	2.246	(1.457,2.915)	2.220	(1.421,2.842)	2.211	(0.958, 4.792)	2.055	(0.796,3.981)
	.9	.1	2.248	(1.235,2.471)	2.236	(1.220,2.441)	2.224	(0.617, 3.083)	2.143	(0.555,2.773)

1. range corresponds to values of $OR_{DF/\overline{E}} \in (0,\infty)$.

TABLE 4.3

Comparison of the Crude Matched Odds Ratio (cORm) With the Uniform, Stratum-Specific Odds Ratio (OR) and the Crude Odds Ratio (cOR): Moderate to Large OR

			$OR_{EF/\overline{D}} = 2$				$OR_{EF/\overline{D}} = 5$			
			$\delta_0 = 0.01$		$\delta_0 = 0.10$		$\delta_0 = 0.01$		$\delta_0 = 0.10$	
OR	v_1	v_0	cORm	cOR[1]	cORm	cOR[1]	cORm	cOR[1]	cORm	cOR[1]
	.1	.9	3.496	(3.320,6.639)	3.471	(3.282,6.564)	3.474	(3.186,15.930)	3.324	(2.966,14.831)
	.3	.7	3.490	(2.963,5.925)	3.429	(2.880,5.760)	3.428	(2.583,12.914)	3.096	(2.167,10.853)
3.5	.5	.5	3.486	(2.610,5.221)	3.411	(2.518,5.035)	3.395	(2.011,10.053)	2.995	(1.597, 7.987)
	.7	.3	3.487	(2.263,4.525)	3.419	(2.189,4.378)	3.384	(1.467, 7.333)	3.022	(1.171, 5.854)
	.9	.1	3.493	(1.920,3.840)	3.462	(1.980,3.780)	3.425	(0.949, 4.747)	3.243	(0.839, 4.197)
	.1	.9	4.991	(4.379,9.478)	4.942	(4.672,9.345)	4.942	(4.532,22.660)	4.651	(4.150,20.750)
	.3	.7	4.977	(4.225,8.450)	4.859	(4.081,8.161)	4.841	(3.647,18.236)	4.229	(2.960,14.800)
5.0	.5	.5	4.969	(3.721,7.441)	4.824	(3.561,7.121)	4.770	(2.825,14.123)	4.063	(2.167,10.838)
	.7	.3	4.970	(3.226,6.451)	4.842	(3.100,6.200)	4.747	(2.085,10.285)	4.129	(1.600, 8.000)
	.9	.1	4.985	(2.740,5.479)	4.926	(2.689,5.378)	4.837	(1.341, 6.705)	4.540	(1.175, 5.875)

1. range corresponds to values of $OR_{DF/\overline{E}} \in (0,\infty)$.

Under general circumstances then, matching can "control" confounding in case-control studies, although never remove it. There is no justification for using cORm over any stratified odds ratio since the bias toward the null implies a less sensitive test of significance based on a crude analysis relative to a stratified one. Indeed, the expected M-H X^2 statistic from a crude analysis is almost always less than the expected M-H X^2 statistic from a stratified analysis. (The question of whether pooling or stratification should accompany matching will be addressed in greater detail in Chapter 5.)

Nevertheless, it may be concluded from the above discussion that matched studies which did not employ stratification should not be condemned as invalid because of probable confounding. While there may remain confounding from variables which were not controlled by the matching or in the analysis, confounding from the matching variables should only be minimal, except for the particular circumstances noted above. Indeed, because of the downward bias associated with the cORm, any adjustment that is necessary will only strengthen positive conclusions and lend support to claims of significant results.

4.4.2 One Matching Variable and Two Confounders

In the previous section the extent to which matching alone can control confounding in follow-up and case-control studies where all of the confounders are matching variables was examined. However, it is not difficult to envisage circumstances in which matching may involve only a few of the confounders which require control. In fact, one might argue that this scenario is in practice quite likely — that

additional variables beyond the matching variables require control in the analysis. If that is the case, the proper analysis of data arising from such a scenario needs to be understood.

Consider the situation where an additional variable beyond the matching variable requires control. Assume that matching is performed on F, a dichotomous extraneous variable, while a second extraneous variable G remains to be controlled. Matching on F alone equates the marginal distributions of F conditional on exposure and non-exposure in a follow-up study (disease and non-disease in a case-control study). Therefore, the expected number of exposed and non-exposed in the i^{th} stratum of F in a follow-up study are

	E	$\overline{E}$
F_i	$N(\theta_{i1} + \theta_{i0})$	$\rho N(\theta_{i1} + \theta_{i0})$

By stratifying on G each of the above expected marginal totals are subdivided according to the distribution of G in that stratum. These distributions are given by

$$P(G_j/F_iE) = \theta_{ij}/(\theta_{i1} + \theta_{i0})$$

$$P(G_j/F_i\overline{E}) = \phi_{ij}/(\phi_{i1} + \phi_{i0})$$

Hence, the expected number of exposed and unexposed subjects in the $(i,j)^{th}$ stratum defined by F and G after matching on the marginal distribution of F are

	E	$\overline{E}$
F_i, G_j	$N\theta_{i1}$	$\rho N\phi_{i1} \frac{(\theta_{i1}+\theta_{i0})}{(\phi_{i1}+\phi_{i0})}$

The expected cell frequencies in the $(i,j)^{th}$ stratum are given below in Table 4.4.

TABLE 4.4

F_i, G_j	E	$\bar{E}$
D	$N\alpha_{ij}\theta_{ij}$	$\rho N\beta_{ij}\phi_{ij}\frac{(\theta_{i1}+\theta_{i0})}{(\phi_{i1}+\phi_{i0})}$
$\bar{D}$	$N(1-\alpha_{ij})\theta_{ij}$	$\rho N(1-\beta_{ij})\phi_{ij}\frac{(\theta_{i1}+\theta_{i0})}{(\phi_{i1}+\phi_{i0})}$
	$N\theta_{ij}$	$\rho N\phi_{i1}\frac{(\theta_{i1}+\theta_{i0})}{(\phi_{i1}+\phi_{i0})}$

A question of interest is whether the analysis may involve data which have been pooled over the matching variable F. It will be shown that if F is associated with the outcome variable in the study, and another variable remains to be controlled in the analysis, then stratification must involve both the matching variable F and the additional variable. In other words both variables will require control in the analysis. Pooling over a matching variable is only allowed if there are no other confounders that require control or if the matching variable is unrelated to the outcome variable. Otherwise, validity will not be preserved.

Consider the analysis of data pooled over F. The expected cell frequencies are those of Table 4.4 summed over the subscript i. In a follow-up study there is no confounding after matching on F if the standardized risk ratio controlling for G (sRR_G) is equal to the standardized risk ratio controlling for F and G ($sRR_{F,G}$) where

$$sRR_G = \frac{\sum_j \alpha_j^* \theta_j^*}{\sum_j \beta_j^* \theta_j^*}$$

and,

$$sRR_{F,G} = \frac{\sum\sum_{ij} \alpha_{ij} \theta_{ij}}{\sum\sum_{ij} \beta_{ij} \theta_{ij}} .$$

The parameters α_j^*, β_j^*, and θ_j^* are defined below:

$$\alpha_j^* = \frac{\alpha_{1j}\theta_{1j} + \alpha_{0j}\theta_{0j}}{\theta_{1j} + \theta_{0j}} ,$$

$$\beta_j^* = \beta_{1j}\phi_{1j}^* + \beta_{0j}\phi_{0j}^* , \quad \text{where} \quad \phi_{ij}^* = \frac{\phi_{ij}\left(\frac{\theta_{i1}+\theta_{i0}}{\phi_{i1}+\phi_{i0}}\right)}{\sum_i \phi_{ij}\left(\frac{\theta_{i1}+\theta_{i0}}{\phi_{i1}+\phi_{i0}}\right)}$$

and,

$$\theta_j^* = \theta_{1j} + \theta_{0j} .$$

Therefore,

$$sRR_G = \frac{\sum_j \left(\frac{\alpha_{1j}\theta_{1j}+\alpha_{0j}\theta_{0j}}{\theta_{ij}+\theta_{0j}}\right)(\theta_{1j}+\theta_{0j})}{\sum_j (\beta_{1j}\phi_{1j}^*+\beta_{0j}\phi_{0j}^*)(\theta_{1j}+\theta_{0j})}$$

$$= \frac{\sum\sum_{ij} \alpha_{ij}\theta_{ij}}{\sum_j \left(\sum_i \beta_{ij}\phi_{ij}^*\right)\left(\sum_i \theta_{ij}\right)}$$

Since the numerators of sRR_G and sRR_{FG} are equal, there is no confounding if

$$\sum_{ij}\sum\beta_{ij}\theta_{ij} = \sum_{j}\left(\sum_{i}\beta_{ij}\phi^{*}_{ij}\right)\left(\sum_{i}\theta_{ij}\right) ,$$

or

$$\sum_{j}\left\{\sum_{i}\beta_{ij}\theta_{ij} - \sum_{i}\beta_{ij}\phi^{*}_{ij}\left(\sum_{i}\theta_{ij}\right)\right\} = 0 ,$$

or equivalently,

$$\sum_{j}\left\{\sum_{i}\beta_{ij}\left[\theta_{ij} - \phi^{*}_{ij}\left(\sum_{i}\theta_{ij}\right)\right]\right\} = 0 . \qquad (4.4.1)$$

The term in brackets can be expressed as

$$+ (\theta_{1j}\phi^{*}_{0j} - \theta_{0j}\phi^{*}_{1j}) , \quad i = 1$$

$$- (\theta_{1j}\phi^{*}_{0j} - \theta_{0j}\phi^{*}_{1j}) , \quad i = 0 .$$

Hence, 4.5.1 reduces to

$$\sum_{j}(\beta_{1j} - \beta_{0j})(\theta_{1j}\phi^{*}_{0j} - \theta_{0j}\phi^{*}_{1j}) = 0 . \qquad (4.4.2)$$

Assuming that F is a risk factor, and F and G do not interact on disease or exposure, then 4.5.2 reduces further to

$$\theta_{1j}\phi^{*}_{0j} = \theta_{0j}\phi^{*}_{1j} = 0 , \quad j = 0,1$$

which can be expressed as

$$\frac{\theta_{1j}\phi^{*}_{0j}}{\theta_{0j}\phi^{*}_{1j}} = 1 , \quad j = 0,1 ,$$

or

$$\frac{\theta_{1j}\phi_{0j}}{\theta_{0j}\phi_{1j}} \cdot \frac{\left(\frac{\theta_{00}+\theta_{01}}{\phi_{00}+\phi_{01}}\right)}{\left(\frac{\theta_{11}+\theta_{10}}{\phi_{11}+\phi_{10}}\right)} = 1 , \quad j = 0,1 ,$$

or

$$\frac{\theta_{1j}\phi_{0j}}{\theta_{0j}\phi_{1j}} = \frac{(\theta_{11}+\theta_{10})(\phi_{01}+\phi_{00})}{(\theta_{01}+\theta_{00})(\phi_{11}+\phi_{10})} , \quad j = 0,1 .$$

Therefore,

$$OR_{EF/G_j} = OR_{EF} , \quad j = 0,1 . \qquad (4.4.3)$$

This result can be argued intuitively. There is no confounding due to F if $OR_{EF/G_j} = 1$, $j = 0,1$. By matching on the marginal distribution of F, OR_{EF} is set to 1 in the data. If (4.4.3) holds, then OR_{EF/G_j} is also set to 1 in the data, and the conditions for no confounding are met. However, it does not matter whether $OR_{EF/G_j} = 1$, $j = 0,1$ unless $OR_{EF} = 1$ as well. If the two odds ratios are not equivalent (and $OR_{EF/G_j} = 1$, $j = 0,1$) in the population, then the matching process will only introduce an association between exposure and F in the data, and confounding will remain uncontrolled. Of course, if F is unrelated to the outcome variable, disease, (F is not a risk factor), then the matching can not introduce confounding and pooling over F is appropriate.

Since it is assumed that G is a confounder, one can demonstrate that (4.4.3) will hold if and only if F and G are not associated. It follows that if G is a confounder, then

$$OR_{EG/F_i} = c , \quad \text{a constant} , \quad i = 0,1 , \quad \text{where} \quad c \neq 1 .$$

Hence,

$$OR_{EF} = \frac{\left(\frac{\phi_{11}\theta_{10}}{\phi_{10}} \cdot c + \theta_{10}\right)(\phi_{01}+\phi_{00})}{\left(\phi_{01}\frac{\theta_{00}}{\phi_{00}} \cdot c + \theta_{00}\right)(\phi_{11}+\phi_{10})}$$

$$= \frac{\frac{\theta_{10}}{\phi_{10}}(\phi_{11} \cdot c + \phi_{10})(\phi_{01}+\phi_{00})}{\frac{\theta_{00}}{\phi_{00}}(\phi_{01} \cdot c + \phi_{00})(\phi_{11}+\phi_{10})}$$

$$= OR_{EF/G_0} \cdot f(c,\underset{\sim}{\phi}) \ .$$

If (4.4.3) holds, then $f(c,\underset{\sim}{\phi}) = 1$, and

$$(\phi_{11}c+\phi_{10})(\phi_{01}+\phi_{00}) = (\phi_{01} \cdot c+\phi_{00})(\phi_{11}+\phi_{10}) \ ,$$

or

$$c\phi_{11}\phi_{01} + c\phi_{11}\phi_{00} + \phi_{10}\phi_{01} + \phi_{10}\phi_{00}$$

$$= c\phi_{01}\phi_{11} + c\phi_{01}\phi_{10} + \phi_{00}\phi_{11} + \phi_{00}\phi_{10} \ ,$$

which can be expressed as

$$\phi_{11}\phi_{00}(c-1) = \phi_{10}\phi_{01}(c-1) \ ,$$

or

$$OR_{FG/\overline{E}} = \frac{\phi_{11}\phi_{00}}{\phi_{01}\phi_{10}} = 1 \ .$$

Also, by the assumption of no interaction, $OR_{FG/E} = 1$.

To summarize, if G is a confounder, then the marginal association between F and exposure is equivalent to the conditional association only if F and G are unrelated. If F and G are related, then (4.4.3) cannot hold, and matching on F will not in and of itself control confounding in a follow-up study.

Consider now the case-control study. Referring to Table 4.4 and replacing the follow-up parameters with the case-control parameters, one can represent the problem in a similar fashion as the follow-up study. In the case-control study where matching on F has been employed, and the data are pooled over F, there is no confounding if

$$sOR_G = sOR_{F,G} \ ,$$

where

$$sOR_G = \left(\sum_j \varepsilon_j^* v_j^*\right) / \sum_j \frac{(1-\varepsilon_j^*)v_j^*\delta_j^*}{(1-\delta_j^*)} ,$$

and

$$sOR_{F,G} = \left(\sum\sum_{ij} \varepsilon_{ij} v_{ij}\right) / \sum\sum_{ij} \frac{(1-\varepsilon_{ij})v_{ij}\delta_{ij}}{(1-\delta_{ij})} .$$

The parameters ε_j^*, δ_j^*, and v_j^* are defined in a similar fashion as α_j^*, β_j^*, and θ_j^*, respectively, from the context of a follow-up study. Therefore,

$$\varepsilon_j^* = \frac{\varepsilon_{1j}v_{1j} + \varepsilon_{0j}v_{0j}}{v_{ij} + v_{0j}}$$

$$\delta_j^* = \delta_{ij}w_{ij}^* + \delta_{0j}w_{0j}^* , \quad w_{ij}^* = \frac{w_{ij}\left(\frac{v_{i1}+v_{i0}}{w_{i1}+w_{i0}}\right)}{\sum_i w_{ij}\left(\frac{v_{i1}+v_{i0}}{w_{i1}+w_{i0}}\right)}$$

$$v_j^* = v_{1j} + v_{0j} .$$

It can be easily shown that the numerator of sOR_G and sOR_{FG} are equal. Comparing denominators, there is no confounding if

$$\sum\sum_{ij}\left(\frac{(1-\varepsilon_{ij})v_{ij}\delta_{ij}}{1-\delta_{ij}}\right) = \sum_j\left(\frac{(1-\varepsilon_j^*)v_j^*\delta_j^*}{1-\delta_j^*}\right)$$

$$= \sum_j \frac{\left(\sum_i (1-\varepsilon_{ij})v_{ij}\right)\left(\sum_i \delta_{ij}w_{ij}^*\right)}{\left(\sum_i (1-\delta_{ij})w_{ij}^*\right)} ,$$

or if

$$\sum_j\left[\sum_i (1-\varepsilon_{ij})v_{ij}\left\{\frac{\delta_{ij}}{1-\delta_{ij}} - \frac{\left(\sum_i \delta_{ij}w_{ij}^*\right)}{\left(\sum_i (1-\delta_{ij})w_{ij}^*\right)}\right\}\right] = 0 ,$$

or

$$\sum_j \left(\sum_j (1-\delta_{ij}) w^*_{ij} \right)^{-1} \left(\sum_i \frac{(1-\varepsilon_{ij}) v_{ij}}{(1-\delta_{ij}) w_{ij}} \left\{ \delta_{ij} w_{ij} \left(\sum_i (1-\delta_{ij}) w^*_{ij} \right) \right. \right.$$

$$\left. \left. - (1-\delta_{ij}) \left(\sum_i \delta_{ij} w^*_{ij} \right) \right\} \right) = 0. \quad (4.4.4)$$

The term in brackets can be reduced to

$$\delta_{ij} w_{ij} \sum_i w^*_{ij} - w_{ij} \sum_i \delta_{ij} w^*_{ij}$$

$$= \begin{cases} w_{ij} w^*_{0j} (\delta_{1j} - \delta_{0j}) , & i = 1 \\ \\ w_{0j} w^*_{1j} (\delta_{0j} - \delta_{1j}) , & i = 0 . \end{cases}$$

Therefore, (4.4.4) can be expressed as

$$\sum_j \left(\sum_i (1-\delta_{ij}) w^*_{ij} \right)^{-1} \left\{ (\delta_{1j} - \delta_{0j}) \left(\frac{(1-\varepsilon_{1j}) v_{ij}}{(1-\delta_{1j}) w_{1j}} \cdot w_{1j} w^*_{0j} \right. \right.$$

$$\left. \left. - \frac{(1-\varepsilon_{0j}) v_{0j}}{(1-\delta_{0j}) w_{0j}} \cdot w_{0j} w^*_{1j} \right) \right\} = 0 . \quad (4.4.5)$$

Assuming no interaction between F and G on exposure and disease, then (4.4.5) holds if either

(a) $\delta_{1j} = \delta_{0j} , \quad j = 0,1 \iff OR_{EF/G_j} = 1 , \quad j = 0,1 , \quad$ or

(b) $\frac{(1-\varepsilon_{1j}) v_{1j}}{(1-\delta_{1j}) w^*_{1j}} \cdot \frac{(1-\delta_{0j}) w^*_{0j}}{(1-\varepsilon_{0j}) v_{0j}} = 1 , \quad j = 0,1 .$

Substituting for w^*_{ij}, expression (b) can also be written as

$$\frac{(1-\varepsilon_{1j}) v_{1j}}{(1-\delta_{1j}) w_{1j}} \frac{(1-\delta_{0j}) w_{0j}}{(1-\varepsilon_{0j}) v_{0j}} \cdot \frac{(v_{01}+v_{00})(w_{11}+w_{10})}{(v_{10}+v_{11})(w_{01}+w_{00})} = 1 , \quad \text{or}$$

$$OR_{DF/\overline{E}G_j} \cdot (OR_{DF})^{-1} = 1 .$$

These results are completely analogous to the follow-up study. If matching has been performed on F and the data are pooled over F, there is no confounding if (a) F is conditionally unrelated to the outcome variable (exposure), or (b) the conditional association between F and disease is equivalent to the unconditional association. Whether F is a risk factor or not has no bearing on the presence or absence of confounding in this context. In fact, if F is not a risk factor, the act of matching may introduce an association between F and disease in the data leaving confounding uncontrolled.

Assuming that the matching variables are indeed risk factors, the nature of confounding will hinge on the exposure / F association. In a case-control study matching does not influence this association in the data, and therefore, will not introduce confounding when there is none ($OR_{EF/G_j} = 1$, $j = 0,1$). In the next chapter the potential loss in efficiency which may result when matching on a non-risk factor will be considered in the context of an additional confounder requiring control in the analysis.

4.5 Validity and Effect Modification

While confounding and effect modification are entirely distinct concepts, the manifestation of one may be directly related to the presence of the other in the data under study. In particular, if the conditions for residual confounding hold, there may be "apparent" effect modification when in fact there is none. Consider the following example for a follow-up study.

Example 4

Let $RR_{ij} = 2$, $\forall (i,j) \Rightarrow \alpha_{ij} = 2 \cdot \beta_{ij}$

(i,j)	β_{ij}	θ_{ij}	ϕ_{ij}
(1,1)	0.001	0.1	0.4
(1,0)	0.00025	0.2	0.3
(0,1)	0.0004	0.4	0.2
(0,0)	0.0001	0.3	0.1

Failing to account for G , $RR_{F_i} = \dfrac{\sum_j \alpha_{ij}\theta_{ij}/(\theta_{i1}+\theta_{i0})}{\sum_j \beta_{ij}\phi_{ij}/(\phi_{i1}+\phi_{i0})}$. Therefore,

$RR_{F_1} = 1.27$, and $RR_{F_0} = 1.81$.

By appearance there is effect modification due to F. In fact, there is no effect modification since the stratum-specific RR's are uniform. Rather, there is confounding due to G which manifests itself as effect modification.

In a follow-up study this circumstance arises in the following way. The assumption of no effect modification requires that

$\alpha_{ij}/\beta_{ij} = RR$, $\forall(i,j)$, where RR = the stratum-specific risk ratio.

However, let us assume that F is accounted for in the analysis while G is not. Then there will be apparent effect modification if $RR_{F_1} \neq RR_{F_0}$. This can be described by three scenarios:

(a) $RR_{F_1} < RR < RR_{F_0}$ (or, $RR_{F_0} < RR < RR_{F_1}$)

(b) $RR \leq RR_{F_1} < RR_{F_0}$ (or, $RR \leq RR_{F_0} < RR_{F_1}$)

(c) $RR_{F_1} < RR_{F_0} \leq RR$ (or, $RR_{F_0} < RR_{F_1} \leq RR$) .

Now,

$$RR_{F_i} \neq RR \Rightarrow \left[\frac{\alpha_{i1}\theta_{i1}+\alpha_{i0}\theta_{i0}}{\alpha_{i1}+\theta_{i0}}\right]\left[\frac{\beta_{i1}\phi_{i1}+\beta_{i0}\phi_{i0}}{\phi_{i1}+\phi_{i0}}\right] \neq RR ,$$

or

$$\frac{RR\beta_{i1}\theta^*_{i1}+RR\beta_{i0}\theta^*_{i0}}{\beta_{i1}\phi^*_{i1}+\beta_{i0}\phi^*_{i0}} \neq RR , \quad \text{where} \begin{cases}\theta^*_{i1} = \theta_{ij}/(\theta_{i1}+\theta_{i0}) \\ \phi_{ij} = \phi_{ij}/(\phi_{i1}+\phi_{i0})\end{cases}$$

and

$$\alpha_{ij} = RR\beta_{ij} , \quad \forall(i,j) .$$

This implies that

$$\beta_{i1}\theta^*_{i1} + \beta_{i0}\theta^*_{i0} \neq \beta_{i1}\phi^*_{i1} + \beta_{i0}\phi^*_{i0} ,$$

and

$$(\beta_{i1}-\beta_{i0})(\theta^*_{i1}-\phi^*_{i1}) \neq 0 , \tag{4.5.1}$$

since

$$\theta^*_{i1} = 1-\theta^*_{i0} , \quad \phi^*_{i1} = 1-\phi^*_{i0} .$$

Therefore, $RR_{F_i} \neq RR$ if

(i) $\beta_{i1} \neq \beta_{i0} \Rightarrow OR_{DG/\bar{E}F_i} \neq 1$

and

(ii) $\theta^*_{i1} \neq \phi^*_{i1} \Rightarrow OR_{EG/F_i} \neq 1$. (4.5.2)

When (i) and (ii) hold for either $i = 0$ or $i = 1$, then there is probable residual confounding due to G. Note that either of (b) or (c)

can result if (4.5.2) holds for one of $i = 0$ or $i = 1$. If (b) and (c) are strict inequalities, then there is indeed residual confounding. Condition (a) describing apparent effect modification may not relate to residual confounding depending upon the nature of the bias at the two levels of F. The biases must be in the same direction if there is no interaction between F and G; hence, condition (a) cannot occur.

In summary, while there is not a one-to-one correspondence between residual confounding and apparent effect modification, a very strong link connects the two concepts. The presence of effect modification should lead an investigator to consider whether a potential confounder, which has not been evaluated, is the cause of the apparent phenomenon before concluding that the variable in question is in fact an effect modifier.

In the case-control study a similar connection between effect modification and confounding can be shown. If there is no effect modification, then $OR_{ij} = OR$, $\forall(i,j)$. Accounting for F in the analysis yields the stratum-specific odds ratio

$$OR_{F_i} = \frac{\left[\frac{\varepsilon_{i1}v_{i1}+\varepsilon_{i0}v_{i0}}{v_{i1}+v_{i0}}\right]\left[1 - \frac{\delta_{i1}w_{i1}+\delta_{i0}w_{i0}}{w_{i1}+w_{i0}}\right]}{\left[\frac{\delta_{i1}w_{i1}+\delta_{i0}w_{i0}}{w_{i1}+w_{i0}}\right]\left[1 - \frac{\varepsilon_{i1}v_{i1}+\varepsilon_{i0}v_{i0}}{v_{i1}+v_{i0}}\right]}$$

$$= \frac{(\varepsilon_{i1}v^*_{i1}+\varepsilon_{i0}v^*_{i0})(1-(\delta_{i1}w^*_{i1}+\delta_{i0}w^*_{i0}))}{(\delta_{i1}w^*_{i1}+\delta_{i0}w^*_{i0})(1-(\varepsilon_{i1}v^*_{i1}+\varepsilon_{i0}v^*_{i0}))}$$

where $v^*_{ij} = v_{ij}/(v_{i1}+v_{i0})$, $w^*_{ij} = w_{ij}/(w_{i1}+w_{i0})$

$$OR_{ij} = OR \ \forall(i,j) \Rightarrow \quad \varepsilon_{ij} = \delta_{ij}OR/(1-\delta_{ij}(1-OR))$$

$$1-\varepsilon_{ij} = (1-\delta_{ij})/(1-\delta_{ij}(1-OR)) \ .$$

Therefore, $OR_{F_i} \neq OR$ implies that

$$\frac{\left[\frac{OR\delta_{i1}\, v^*_{i1}}{1-\delta_{i1}(1-OR)} + \frac{OR\delta_{i0}\, v^*_{i0}}{1-\delta_{i0}(1-OR)}\right]((1-\delta_{i1})w^*_{i1}(1-\delta_{i0})w^*_{i0})}{(\delta_{i1}w^*_{i1}+\delta_{i0}w^*_{i0})\left[\frac{(1-\delta_{i1})v^*_{i1}}{1-\delta_{i1}(1-OR)} + \frac{(1-\delta_{i0})v^*_{i0}}{1-\delta_{i}(1-OR)}\right]} \neq OR \ ,$$

or,

$$\left[\frac{\delta_{i1}v^*_{i1}}{1-\delta_{i1}(1-OR)} + \frac{\delta_{i0}v^*_{i0}}{1-\delta_{i0}(1-OR)}\right](1-(\delta_{i1}w^*_{i1}+\delta_{i0}w^*_{i0}))$$

$$\neq \left[\frac{v^*_{i1} - \delta_{i1}v^*_{i1}}{1-\delta_{i1}(1-OR)} + \frac{v^*_{i0} - \delta_{i0}v^*_{i0}}{1-\delta_{i0}(1-OR)}\right](\delta_{i1}w^*_{i1}+\delta_{i0}w^*_{i0}) \ ,$$

$$\frac{\delta_{i1}v^*_{i1}}{1-\delta_{i1}(1-OR)} + \frac{\delta_{i0}v^*_{i0}}{1-\delta_{i0}(1-OR)} \neq \left[\frac{v^*_{i1}}{1-\delta_{i1}(1-OR)} + \frac{v^*_{i0}}{1-\delta_{i0}(1-OR)}\right](\delta_{i1}w^*_{i1}+\delta_{i0}w^*_{i0}) \ ,$$

and

$$\frac{v^*_{i1}}{1-\delta_{i1}(1-OR)}(\delta_{i1}-(\delta_{i1}w^*_{i1}+\delta_{i0}w^*_{i0})) + \frac{v^*_{i0}}{1-\delta_{i0}(1-OR)}(\delta_{i0}-(\delta_{i1}w^*_{i1}+\delta_{i0}w^*_{i0})) \neq 0.$$

And, since $w^*_{i1} = 1 - w^*_{i0}$,

$$\frac{v^*_{i1}w^*_{i0}}{1-\delta_{i1}(1-OR)}(\delta_{i1}-\delta_{i0}) + \frac{v^*_{i0}w^*_{i1}}{1-\delta_{i0}(1-OR)}(\delta_{i0}-\delta_{i1}) \neq 0 \ .$$

Then,

$$(\delta_{i1}-\delta_{i0})\left[\frac{v^*_{i1}w^*_{i0}}{1-\delta_{i1}(1-OR)} - \frac{v^*_{i0}w^*_{i1}}{1-\delta_{i0}(1-OR)}\right] \neq 0 \ .$$

Also, $1 - \delta_{ij}(1-OR) = (1-\delta_{ij})OR/(1-\varepsilon_{ij})$, therefore,

$$(\delta_{i1}-\delta_{i0})\left[\frac{v^*_{i1}w^*_{i0}(1-\varepsilon_{i1})}{(1-\delta_{i1})OR} - \frac{v^*_{i0}w^*_{i1}(1-\varepsilon_{i0})}{(1-\delta_{i0})OR}\right] \neq 0 \ .$$

Hence, both

(a) $\delta_{i1} \neq \delta_{i0} \Rightarrow OR_{EG/DF_i} \neq 1$

and

(b) $\dfrac{v^*_{i1}w^*_{i0}(1-\varepsilon_{i1})}{(1-\delta_{i1})} \neq \dfrac{v^*_{i0}w^*_{i1}(1-\varepsilon_{10})}{(1-\delta_{i0})}$, or

$$OR_{DG/\overline{E}F_i} = \frac{(1-\varepsilon_{i1})v^*_{i1}(1-\delta_{i0})w^*_{i0}}{(1-\delta_{i1})w^*_{i1}(1-\varepsilon_{i0})v^*_{i0}} \neq 1$$

must hold for $OR_{F_i} \neq OR$.

It is clear that the scenario regarding effect modification and residual confounding described earlier for the follow-up study applies equally to the case-control study. If effect modification is observed, failing to take into account the variable G, then the phenomenon may have resulted because G is actually a confounder.

CHAPTER 5

THE RELATIVE EFFICIENCY OF MATCHING AND RANDOM SAMPLING: TWO POTENTIAL CONFOUNDING VARIABLES

5.1 Introduction

In this chapter the results of Chapter 3 will be extended to two potential confounding variables. The introduction of a second extraneous variable into the analysis allows a number of issues concerning the relative efficiency of matching and random sampling to be considered. For example, two particular concerns relative to the question of the efficiency of matching are (a) whether "overmatching," or matching on variables which are not confounders, will lead to a loss in efficiency with respect to not matching on those variables; and (b) whether matching on variables which are correlated might be less efficient than matching on simply one of the variables or not matching at all. These two issues are dealt with in Sections 5.4 and 5.5. Two important questions which were raised in Chapter 3 and are addressed again in this chapter are the characterization of the relationship between the nature of the (underlying) confounding and the relative efficiency of matching to random sampling, as well as the effect of loss of sample size. Sections 5.5 and 5.6 are devoted to these considerations. Another topic which is treated in this chapter is the optimal analyses from the point of view of efficiency (in terms of the M-H X^2

statistic) of matched data and randomly sampled data when there is no confounding.

Evaluation of the relative efficiency of matching to random sampling is complicated by involving a second confounder. For this reason the focus of attention of this chapter is somewhat restricted. Only one value of a local alternative to the null measure of effect is considered (RR, OR = 1.5). The analyses are restricted to the conditions of "no interaction" between the two confounders, F and G, on disease and on exposure (Section 4.2.4). This restriction is followed for the sake of simplifying the analyses, even though it does not necessarily represent what is found in practice. Also, the "baseline" disease and exposure probabilities, denoted by β_{00} and δ_{00}, are restricted to 0.0001 and 0.05, respectively, in order to provide representative parameter values for each of the follow-up and case-control studies. The value $\delta_{00} = 0.05$ implies large exposure probabilities for the case-control study relative to the disease probabilities of the follow-up study, the major contrast between the underlying models of the two studies.

The values for the relative efficiency reported in this chapter are <u>averages</u> of individual asymptotic relative efficiences (ARE's), defined in Chapter 3. Averaging is conducted over a spectrum of values of $\underset{\sim}{\theta}$ and $\underset{\sim}{\phi}$ in the follow-up study ($\underset{\sim}{\gamma}$ and $\underset{\sim}{\psi}$ in the case-control study) for which odds ratios representing the associations between each confounder and disease and exposure, and the intercorrelation between the confounders are specified (Section 4.2.3). In addition, the restrictions $\sum\sum_{ij}\theta_{ij} = \sum\sum_{ij}\phi_{ij} = 1$, and "no interaction" (Section 4.2.4) are followed). The above restrictions, after introducing β_{00} and RR,

determine all of the disease probabilities in a follow-up study (similarly for the case-control study). Otherwise, the restrictions and specifications allow two parameters, say θ_{11} and θ_{10} (v_{11} and v_{10}), to vary. For the purposes of this study they were allowed to vary between 0.05 and 0.85 in increments of 0.10, as long as all of the other θ_{ij} and ϕ_{ij} (v_{ij} and w_{ij}) also fall between 0.05 and 0.85. In this way a reasonable number of values for the relative efficiency was generated for each unique set of specifications of the aforementioned odds ratios. The usual number of individual values generated was between 10 and 25, and the range of values was usually less than 10%. Hence, the average ARE's displayed in this chapter reflect the relative efficiency one would expect to find for specific values of the odds ratios relating to the nature of the underlying, potential confounding in the population under study. They should not be interpreted as true ARE's but rather as indices of relative efficiency.

5.2 Optimal Analyses of Matched Data

When matching is employed to choose the referents in either a follow-up or case-control study, the decision to stratify the data in constructing the M-H X^2 statistic is not obvious. Birch (1964) has shown that the M-H X^2 statistic is a function of the stratum-specific log odds ratios and is optimal when the odds ratio is uniform across the strata. Since matching does not remove confounding when the odds ratio is the measure of effect, it would seem that to attain validity, stratification on the matching variables should always be employed.

However, for many practical situations it has been shown that the negative bias resulting from a crude analysis of matched data is negligible. (See the discussion in Chapter 4.) There may therefore be an advantage with respect to efficiency to pool the data rather than stratify, while tolerating a small bias in the estimated odds ratio.

In order to test that hypothesis, M-H X^2 statistics constructed from expected cell frequencies (based on matching), which had been pooled and stratified were contrasted. Let X_p^2 and X_s^2 represent the "expected" M-H statistics under pooling and stratification, respectively. In the context of the follow-up study, the expected stratum-specific cell frequencies are given by Table 1.4, where ρN_1 is substituted for N_0. Therefore,

$$X_p^2 = \frac{\left[\frac{\left(\sum_i N_1\alpha_i\theta_i\right)\left(\sum_i \rho N_1(1-\beta_i)\theta_i\right) - \left(\sum_i \rho N_1\beta_i\theta_i\right)\left(\sum_i N_1(1-\alpha_i)\theta_i\right)}{N_1(1+\rho)}\right]^2}{\frac{N_1\rho N\cdot\left(\sum_i N_1\theta_i(\alpha_i+\rho\beta_i)\right)\left(N_1(1+\rho)-\sum_i N_1\theta_i(\alpha_i+\rho\beta_i)\right)}{(N_1(1+\rho))^2(N_1(1+\rho)-1)}}$$

$$= \frac{(1+\rho)\rho\left[\sum_i \theta_i(\alpha_i-\beta_i)\right]^2}{\frac{\left[\sum_i \theta_i(\alpha_i+\rho\beta_i)\right]\left[(1+\rho)-\sum_i \theta_i(\alpha_i+\rho\beta_i)\right]}{N_1-(1+p)^{-1}}} \qquad (5.2.1)$$

and

$$X_s^2 = \frac{\left[\sum_i \frac{(N_1\alpha_i\theta_i)(\rho N_1(1-\beta_i)\theta_i)-(\rho N_1\beta_i\theta_i)(N_1(1-\alpha_i)\theta_i)}{N_1(1+\rho)\theta_i}\right]^2}{\sum_i \frac{N_1\theta_i\ \rho N_1\theta_i(N_1\theta_i(\alpha_i+\rho\beta_i))\ (N_1(1+\rho)\theta_i - N_1\theta_i(\alpha_i+\rho\beta_i))}{(N_1(1+\rho)\theta_i)^2(N_1(1+\rho)\theta_i-1)}}$$

$$= \frac{(1+\rho)\rho\left[\sum_i \theta_i(\alpha_i-\beta_i)\right]^2}{\sum_i \frac{\theta_i^2(\alpha_i+\rho\beta_i)((1+\rho)-(\alpha_i+\rho\beta_i))}{N_1\theta_i-(1+\rho)^{-1}}} . \qquad (5.2.2)$$

The two statistics have been expressed in such a way that the numerators are equal. Let D_p and D_s be the denominators of X_p^2 and X_s^2, respectively. Then

$$D_p - D_s = \frac{(1+\rho)}{N_1-(1+\rho)^{-1}}\sum_i\theta_i(\alpha_i+\rho\beta_i) - \frac{\left[\sum_i\theta_i(\alpha_i+\rho\beta_i)\right]^2}{N_1-(1+\rho)^{-1}}$$

$$- (1+\rho)\sum_i \frac{\theta_i^2(\alpha_i+\rho\beta_i)}{N_1\theta_i-(1+\rho)^{-1}} + \sum_i \frac{\theta_i^2(\alpha_i+\rho\beta_i)^2}{N_1\theta_i-(1+\rho)^{-1}} \qquad (5.2.3)$$

$$= \frac{(1+\rho)}{N_1-(1+\rho)^{-1}}\sum_i\Delta_i - \frac{\left(\sum_i\Delta_i\right)^2}{N_1-(1+\rho)^{-1}} - (1+\rho)\sum_i \frac{\theta_i\Delta_i}{N_1\theta_i-(1+\rho)^{-1}}$$

$$+ \sum_i \frac{\Delta_i^2}{N_1\theta_i-(1+\rho)^{-1}}, \quad \text{where } \Delta_i = \theta_i(\alpha_i+\rho\beta_i)$$

$$= \frac{(1+\rho)N_1}{N_1-(1+\rho)^{-1}}\sum_i\Delta_i(1-N_1\Delta_i)\left(\frac{\theta_i-1}{N_1\theta_i-(1+\rho)^{-1}}\right) - 2\sum_{i<j}\sum \frac{\Delta_i\Delta_j}{N_1-(1+\rho)^{-1}} . \qquad (5.2.4)$$

Note that the $N_1\Delta_i$ represent the expected number of diseased subjects in the i^{th} stratum. Therefore, unless the data have been spread over too many strata, $N_1\Delta_i$ should generally exceed 2. The first term of

(5.2.4) is therefore positive, since $\theta_i < 1$, $\forall$ i. The second term of (5.2.4) is also positive; the sign of (5.2.4) is then undetermined.

Numerical evaluations of D_s/D_p were conducted at a variety of of values of the parameters for both studies in order to compare the size of D_s relative to D_p (and, therefore, the size of X_p^2 relative to X_s^2). Table 5.1 displays a selection of averages of D_s/D_p in the context of two dichotomous confounders, F and G.

The results indicate that D_s/D_p is almost without exception less than 1. When the average ratio exceeds 1, it does so by no more than 0.003. However, when D_s/D_p is less than 1, espcially when F and G are strong potential confounders, the loss in efficiency due to pooling can exceed 10%, particularly for the case-control study.

In most circumstances stratification should accompany matching in the analysis. While stratification is employed to preserve validity, it is now clear that it also results in a more efficient analysis. A possible explanation is that a pooled M-H analysis corresponds to an estimated odds ratio, cORm, which is biased toward the null, However, a more likely answer is that controlling for a variable which is associated with the response (disease in a follow-up study; exposure in a case-control study) tends to increase the efficiency, which is the same motivation for the analysis of covariance.

For the case of no confounding the appropriate method of analysis depends on the study type. If the matching variables are risk factors, one should stratify on the matching variables in a follow-up study. Since matching controls confounding in a follow-up study, whether or not there is underlying confounding has no bearing on the choice of

analysis. If the matching variables in a case-control study are also risk factors for the disease, there is no confounding if $\delta_{ij} = \delta$, $\forall$ (i,j), or if $OR_{EF/DG_j} = OR_{EG/DF_i} = 1$, $\forall$ i and j. If, in addition, uniformity of the odds ration is assumed, then $\varepsilon_{ij} = \varepsilon$, $\forall$ (i,j), Hence, (5.2.3) becomes (in terms of the case-control parameters):

$$D_p - D_s = \frac{(1+\rho)}{N_1-(1+\rho)^{-1}}(\varepsilon+\rho\delta)\sum_i v_i - \frac{(\varepsilon+\rho\delta)^2}{N_1-(1+\rho)^{-1}}\sum_i v_i$$

$$- (1+\rho)(\varepsilon+\rho\delta)\sum_i \frac{v_i}{N_1 v_i-(1+\rho)^{-1}} + (\varepsilon+\rho\delta)^2\sum_i \frac{v_i^2}{N_1 v_i(1+\rho)^{-1}}$$

$$= \left[(1+\rho)(\varepsilon+\rho\delta)-(\varepsilon+\rho\delta)^2\right]\left[\frac{1}{N_1-(1+\rho)^{-1}} - \sum_i \frac{v_i^2}{N_1 v_1-(1+\rho)^{-1}}\right]$$

$$= \frac{\left[(1+\rho)(\varepsilon+\rho\delta)-(\varepsilon+\rho\delta)^2\right]}{N_1-(1+\rho)^{-1}}\left[1 - \sum_i \frac{v_i^2(N_1-(1+\rho)^{-1})}{N_1 v_i-(1+\rho)^{-1}}\right].$$

The second term in brackets can also be written as

$$1 - \sum_i v_i \frac{(N_1-(1+\rho)^{-1})}{(N_1-(v_i(1+\rho))^{-1}}$$

which is always negative since $N_1 - (1+\rho)^{-1} > N_1 - (v_i(1+\rho))^{-1}$, $\forall$ i. The first term in brackets is always positive, as

$$(\varepsilon+\rho\delta) < 1+\rho, \quad \forall\ \varepsilon,\ \delta,\ \text{and}\ \rho .$$

Therefore, in a case-control study it is more efficient to pool over matching variables which are non-confounders (independent of exposure). For the case of two dichotomous matching variables, the gain in efficiency is so small (about 1%) that the choice to pool or not is

rather academic. Nevertheless, as the number of matching variables increases (and the corresponding number of strata), the loss in efficiency from stratifying may increase.

5.3 No Confounding

5.3.1 Follow-up Study

Assuming that variables for which matching is to be considered are risk factors for the disease, the conditions for no confounding are given by $OR_{EF/G_j} = 1$ and $OR_{EG/F_i} = 1$, or by (4.2.2b):

$$\theta_{ij} = \phi_{ij} , \quad \forall (i,j) .$$

The process of matching replaces the ϕ_{ij} with θ_{ij} so that the tables of expected cell frequencies for matching and random sampling are identical. Clearly, the relative efficiency of matching to random sampling is 1, assuming that the same analysis is applied to the data from each design. The results of Section 5.2 indicate that the more efficient analysis involves stratification. Samuels (1980) claims that matching under these conditions increases efficiency, but only because stratification is not advocated for random sampling as it is for matching. In reality, matching is a futile endeavor which neither results in a gain or loss in efficiency (Kupper, et al., 1980).

5.3.2 Case-Control Study

The conditions for no confounding in a case-control study, where the matching variables are assumed to be risk ractors, are given by

(4.2.4a):

$$\delta_{ij} = \delta \ , \quad \forall \ (i,j) \ .$$

The expected M-H X^2 statistic from matching after pooling over F and G is given by (5.2.1), which can be written as

$$X_m^2 = \frac{(1+\rho)\rho(\varepsilon-\delta)^2\left[\sum_i \sum_j v_{ij}\right]^2}{\left[(\varepsilon+\rho\delta)\sum_i \sum_j v_{ij}\right]\left[(1+\rho)-(\varepsilon+\rho\delta)\sum_i \sum_j v_{ij}\right]/(N_1-(1+\rho)^{-1})}$$

$$= \frac{\rho(N_1-(1+\rho)^{-1})(\varepsilon-\delta)^2}{(\varepsilon+\rho\delta)[(1+\rho)-(\varepsilon+\rho\delta)]} \ . \qquad (5.3.1)$$

Digressing temporarily from the issue of relative efficiency, the question of whether randomly sampled data should be pooled or stratified in the analysis, if there is no confounding, might be raised here. The expected M-H X^2 statistic based on pooled data, assuming no confounding and uniformity of the odds ratio, is given by:

$$X_p^2 = \frac{\left[\frac{\left(\sum\sum_{ij} N_1 \varepsilon_{ij} v_{ij}\right)\left(\sum\sum_{ij} \rho N_1 (1-\delta_{ij}) w_{ij}\right) - \left(\sum\sum_{ij} N_1 (1-\varepsilon_{ij}) v_{ij}\right)\left(\sum\sum_{ij} \rho N_1 \delta_{ij} w_{ij}\right)}{N_1(1+\rho)}\right]^2}{\frac{N_1 \cdot \rho N_1 \left(\sum\sum_{ij} N_1 (\varepsilon_{ij} v_{ij} + \rho\delta_{ij} w_{ij})\right)\left(N_1(1+\rho) - \sum\sum_{ij} N_1 (\varepsilon_{ij} v_{ij} + \rho\delta_{ij} w_{ij})\right)}{(N_1(1+\rho))^2 (N_1(1+\rho)-1)}}$$

$$= \frac{(N_1-(1+\rho)^{-1})(\varepsilon-\delta)^2}{(\varepsilon+\rho\delta)(1+\rho-(\varepsilon+\rho\delta))} \ . \qquad (5.3.2)$$

The expected M-H X^2 statistic based on stratification is taken from (3.3.2), where

$$X_s^2 = \frac{\rho(\varepsilon-\delta)^2\left[\sum\sum_{ij}\frac{v_{ij}w_{ij}}{v_{ij}+\rho w_{ij}}\right]^2}{\sum\sum_{ij}\frac{v_{ij}w_{ij}(\varepsilon_{ij}v_{ij}+\rho\delta_{ij}w_{ij})(v_{ij}+\rho w_{ij}-(\varepsilon_{ij}v_{ij}+\rho\delta_{ij}w_{ij}))}{(v_{ij}+\rho w_{ij})^2(N_1(v_{ij}+\rho w_{ij})-1)}}.$$

The expression $X_p^2 - X_s^2$ has no analytic solution. Hence, the ratio X_p^2/X_s^2 was evaluated at various levels of the parameters in order to describe numerically the effect of pooling on the efficiency. Table 5.2 summarizes the results for the set of values: OR = 1.5, $OR_{FG/\bar{D}} = 2$, $\delta_{00} = 0.05$, and $\rho = 1$. These conditions can be considered typical of a wide range of situations in which a case-control study might be implemented.

The failure to pool can clearly lead to a substantial loss in efficiency if the extraneous variables are strong risk factors. As opposed to the follow-up study, it is essential that the analysis of randomly sampled data be applied to pooled rather than stratified data in the case-control setting, under the conditions of no confounding.

Returning to the issue of relative efficiency, the two methods are seen to be equally efficient. Following the admonition to pool, (5.3.2) is equivalent to (5.3.1), and the relative efficiency is 1. In a case-control study if the matching involves risk factors, and there is no confounding, both matching and random sampling provide equal efficiency. These results agree in principle with those of Samuels (1979).

5.4 Residual Confounding

In this context a particularly important issue regarding matching arises: whether "overmatching," or matching on a non-confounder, will lead to a loss in efficiency when there are additional confounding variables which require control. In Section 5.3 the issue of overmatching was studied under the assumption that matching involved all of the potential confounding variables. The results indicated that for both studies there was no loss in efficiency when matching on non-confounders.

As discussed in Chapter 4, there is no reason to expect that in practice all of the confounding variables will be included in the matching process. After the subjects have been selected, new confounders may emerge from inspection of the data (assuming the data include information on these variables), and control for these variables would be initiated at the analysis stage. The issue of overmatching will be considered in this context.

Let us assume that of two potential confounding variables, F and G, G is a confounder but F is not. There are four possible sampling schemes which need to be evaluated under this assumption: (a) matching on F and G, (b) matching on F only, (c) matching on G only, and (d) random sampling. In Chapter 3 sampling schemes (c) and (d) were evaluated relative to one another, and will not be discussed here. Methods (a) and (b) represent overmatching on the non-confounder F. In this section matching on F and G as well as matching on F only will be evaluated relative to random sampling.

5.4.1 Follow-up Study

The expected X^2 statistic from matching on F when there is residual confounding due to G can be expressed by substituting Δ_{ij} for ϕ_{ij} in (3.3.2), where

$$\Delta_{ij} = \phi_{ij} \frac{(\theta_{ij}+\theta_{i0})}{(\phi_{i1}+\phi_{i0})} .$$

In Chapter 4 it was demonstrated that validity generally cannot be preserved unless the data are stratified on both F and G. The expected M-H X^2 statistic for matching on F and G is given by (5.2.2), where the data are stratified on both F and G. The expected M-H X^2 statistic for random sampling (3.3.2) is based upon stratification on G only.

Evaluations of each matching scheme vs. random sampling are displayed in Table 5.3. While limited to specific values of β_{00} and ρ, these results are comparable to more general conditions where local alternatives to the null are involved. As a reminder, the entries in this table are averages of individual asymptotic relative efficiencies, where averaging is conducted over a range of values of $\underset{\sim}{\theta}$ and $\underset{\sim}{\phi}$ such that certain restrictions and parameter specifications are met. (See discussion in Section 5.1.)

With respect to the potential loss in efficiency from overmatching in a follow-up study, the results in Table 5.3 support the following conclusion. If the confounder G is not included in the matching, then matching on F has little effect on the efficiency relative to random sampling. Only if G is a strong confounder, F is a strong risk factor, and F and G are highly related will there be a

substantial gain or loss by matching on the non-confounder F. In the situation where there is a large loss in efficiency by matching on F, G is a strong "negative" confounder and F is highly related to G, or G is a strong "positive" confounder and there is a strong inverse relationship between F and G. Hence, matching on a correlate of a "negative" confounder seems to cause a loss in efficiency even though the matching variable is itself a non-confounder. Matching on a non-confounder is only useful if that variable is a strong risk factor and is directly related to a strong "positive" confounder (or inversely related to a strong "negative confounder). This point will be pursued further in Section 5.5.

5.4.2 Case-Control Study

Matching on a non-confounder ($OR_{EF/\bar{D}G_j} = 1$, $j = 0,1$) in a case-control study leads to quite similar results with respect to efficiency as in the follow-up study. Table 5.4 summarizes the results of numerical evaluations of the relative efficiency for both matching on F only and matching on F and G. These results confirm the notion that matching on a non-confounder may lead to a moderate gain in efficiency over random sampling if that variable is directly related to a confounder, especially if the relationship is strong. For the case-control study this appears to be true regardless if the confounder is "negative" or "positive."

Other conclusions outlined in Section 5.4.1 apply equally to the context of the case-control study.

5.5 Joint Confounding

For the case of joint confounding two questions are addressed in this section. First, the manner in which the relative efficiency is related to the strength and direction of the confounding potential of each variable, as well as their interdependence, is investigated. Secondly, the question of whether the relative efficiency is affected by the failure to match on one of two confounders is considered. This work contrasts with that of the previous section, where the same question was posed under the assumption that the matching involved a non-confounder.

In this section matching on F only and matching on F and G, where both F and G are confounders, are again compared to random sampling. All three expected M-H X^2 statistics are stratified on both F and G. The results in this section are again limited to a single value of the measure of effect, which represents a local alternative to the null. Other restrictions outlined in the introduction to this chapter are also followed.

The results of numerical evaluation of averaged ARE's are displayed in Table 5.5 to 5.16. Where there are no entries in the tables, no values of $\underset{\sim}{\theta}$ or $\underset{\sim}{\phi}$ could be generated in which all of the specifications and restrictions could be met.

The results of this section are difficult to summarize because the nature and direction of confounding for multiple confounders is difficult to describe. Recall that in Chapter 4 the conditions for no joint confounding could not be written in terms of measures of association

between the confounders and disease and exposure. For that reason the direction of the joint confounding was studied numerically for each entry of Tables 5.5 to 5.16. Each of these tables is structured in the format of Figure 5.1.

Figure 5.1 represents the region which characterizes the associations between the confounders F and G and exposure in a follow-up study (or disease in a case-control study). By arbitrarily choosing the associations between each confounder and the outcome variable to be direct (positive), all possible representations of confounding involving two confounders can be described by the regions in Figure 5.1. Hence, Region I represents the case where F and G are both "negative" confounders. Regions II and III represent the cases where one variable is a "negative" confounder, and the other is a "positive" confounder. And, in Region IV both variables are "positive" confounders. Recall that "positive" confounding is an upward bias in the measure of effect, and "negative" confounding is a downward bias, or a bias toward the null. A "positive" confounder, if left uncontrolled, causes "positive" confounding, etc.

Numerical evaluations in this chapter imply that Regions I and IV represent joint "negative" and "positive" confounding, respectively, although it is conceivable on the basis of (4.2.26) that there could be exceptions to the rule. The direction of confounding in Regions II and III can be either positive or negative or there can be no confounding, which reveals the equivocal nature of confounding when multiple variables are considered. The relative efficiency shall be discussed with reference to the regions in Figure 5.1.

5.5.1 Follow-up Study

The results for the follow-up study are displayed in Table 5.5 to 5.10. A number of general conclusions can be drawn regarding matching.

(a) Matching is likely to be more efficient than random sampling, whether the matching incorporates all or only a few of the confounding variables. This is true especially for situations where the confounding is mild to moderate (Tables 5.5 and 5.8). In addition, the expected gain from matching outweighs the expected loss in efficiency. Only in Region I (joint "negative" confounding) can there be an appreciable loss in efficiency from matching. These conclusions parallel those of Chapter 3.

(b) There is a significant relationship between the direction of the confounding associated with a potential matching variable, and the relative merits of matching on that variable. Generally, there is no gain in efficiency from matching on a "negative" confounder unless that variable is a strong confounder. Consider Regions II and IV of any of Tables 5.5 to 5.10. In each case matching on F and G is more efficient than matching on F because G is a "positive" confounder. In contrast are Regions I and III, where in many instances matching on F only is at least as efficient as matching on both F and G. (Note especially Tables 5.6 and 5.9.) In addition, the lack of contrast between Regions III IV of Table 5.6 is revealing. As a group the relative efficiencies in Region III are only slightly lower than those of

Region IV even though in Region III G is a "negative" confounder. The variable F is a strong positive confounder; however, which seems to dominate the weaker variable G. The key element to gaining efficiency by matching is to match on strong "positive" confounders.

(c) It is generally more efficient to match on independent confounders (Tables 5.5, 5.6, 5.7) rather than correlated confounders (Tables 5.8, 5.9, 5.10). As F and G are increasingly interdependent, the relative increase in efficiency of matching over random sampling occurs only when both variables are "positive" confounders (Region IV). Otherwise, the relative efficiency is not enhanced by matching on variables which are highly related, especially if both are "negative" confounders. As an example consider the contrast of Tables 5.7 and 5.10. An implication of this conclusion is that unless F and G are both "positive" confounders, one should expect a lesser gain in efficiency by matching on highly related confounders than by matching on independent confounders.

5.5.2 Case-Control Study

The results of numerical evaluations of the ARE's for the case-control study are quite consistent with the findings of the follow-up study and of the results of Chapter 3. Tables 5.11 to 5.16 display the averaged relative efficiencies for OR = 1.5, $\delta_{00} = 0.05$, $\rho = 1$, and specific levels of association between F and G, and exposure and disease. Some major distinctions between the results of the two studies include the following:

(a) Matching in a case-control study has almost universal superiority over random sampling in terms of efficiency. Inspection of Tables 5.10 to 5.16 reveals that regardless of the nature of potential confounding associated with each matching variable, the strength of dependence between F and G, or whether the matching has included both confounders or not, matching is likely to yield a larger M-H X^2 statistic than random sampling. At the same time matching does not yield as sizable a gain in efficiency in a case-control study as in a follow-up study. The conclusions confirm those in Chapter 3.

(b) Following from the comments in (a) matching on a "negative" confounder is less likely to result in a loss in efficiency. In fact, matching leads to a gain in efficiency in some areas of Region of every table presented in this section (5.10 to 5.16). There still are losses in this region but they are minimal in comparison to the gains which are achieved there. For the case-control study the stronger the confounding (regardless of direction), the greater the gain in efficiency from matching over random sampling.

(c) The strength of the association between F and G is of lesser importance with regard to the relative efficiency. There can, however, be a substantial gain in efficiency (as opposed to a loss in efficiency in a follow-up study) when two highly related, "negative" confounders are the matching variables. As an example of this phenomenon, see Region I of Tables 5.14, 5.15, and 5.16.

5.6 Loss of Sample Size and Relative Efficiency

In Chapter 3 it was demonstrated that the most damaging effect on the efficiency of matching relative to random sampling is the loss in sample size which can occur in the process of selecting subjects by matching. If the matching process is assumed to discard subjects which would otherwise be available in a non-matched study sample, then the expected loss in efficiency was shown to be enough override any advantage of matching. The results of this section support that claim as well.

Numerical evaluations were conducted similar to those of Section 5.5 while incorporating Method B quantification of loss (3.4.1). The results were overwhelmingly in favor of random sampling. Table 5.17 displays a typical set of circumstances from which a matched sample, assuming loss of sample subjects, might arise. The entries in Table 5.17 are, again, averages of ARE's generated under that specific set of circumstances. (One should note that the variation in individual relative efficiencies was somewhat greater than the variations experienced for the relative efficiencies in Section 5.5. A probable explanation is the dependence of loss on $\min_i(\phi_i/\theta_i)$, which can vary for a given set of odds ratios measuring confounding, etc.). Table 5.17 should be compared to Table 5.11, representing the relative efficiency without incorporating loss of sample size and also Table 3.14 in Chapter 3.

Almost without exception the relative efficiency is likely to be less than 0.5θ unless the underlying confounding is mild. The nature and strength of confounding are of relatively little consequence on

the relative efficiency when loss of sample size is assumed. Comparing Table 5.17 to Table 3.14 suggests that the relative efficiency decreases as additional confounding variables are included in the matching, perhaps because $\min_i(\phi_i/\theta_i)$ decreases as the range of i increases. Regardless, under no circumstances should matching be used as the method of subject selection if there will be a loss of sample size from matching, as modelled by the method used here and in Chapter 3.

5.7 Summary

The results of this chapter can be summarized by the following statements. First, the most important concern with regard to matching is whether there will be a loss of sample size from matching. If so, matching should not be the method of subject selection. Second, the most important characteristic of a potential matching variable is whether it is a "positive" or "negative" confounder, or if not a confounder, whether it is strongly related to another confounder. For the follow-up study a matching variable which is either a "negative" confounder or strongly related to one will not generally lead to an increase in efficiency if selected for matching. On the other hand a "positive" confounder (or a variable highly related to one), will lead to a sizeable gain in efficiency if incorporated in the matching. For the case-control study matching on any type of confounder will quite likely lead to a gain in efficiency.

Third, matching on a non-confounder in general has no appreciable effect on the efficiency relative to random sampling unless it is highly related to another confounder. That is, overmatching is not a problem

with regard to the relative efficiency of matching with respect to random sampling.

Fourth, from a general perspective matching is a useful method to increase the efficiency of a test of association between disease and exposure in an epidemiologic study. The risk of losing efficiency by matching is offset by the sizeable gains which may result; therefore, matching should be recommended as a method of subject selection for epidemiologic studies.

FIGURE 5.1

Direction of Confounding[1] and Types of Confounding Variables

$OR^2_{EF/G}$ $(OR_{DF/\bar{E}G})^3$	$OR_{EG/F}$ [2] $(OR_{DG/\bar{E}F})$ [3] < 1	$OR_{EG/F}$ [2] $(OR_{DG/\bar{E}F})$ [3] > 1
< 1	I F+G "Negative" Confounders Joint "Negative" Confounding	II F "Negative" Confounder G "Positive" Confounder Nature of Confounding Undetermined
> 1	III F "Positive" Confounder G "Negative" Confounder Nature of Confounding Undetermined	IV F+G "Positive" Confounder Joint "Positive" Confounding

1. The associations between each confounder and the outcome variables are assumed to be direct (the corresponding odds ratios are greater than 1) without loss of generality.

2. Follow-up Study

3. Case-Control Study

TABLE 5.1

Average Relative Efficiency of Pooling and Stratification in the Analysis of Matched Data

Follow-up Study (RR = 1.5, ρ = 1)		Values of ($OR_{DF/\bar{E}}$, $OR_{DG/\bar{E}F}$)		
β_{00}	$OR_{FG/E}$	(2,2)	(2,5)	(5,5)
0.0001	0.5	1.002	1.002	1.001
	5.0	1.002	1.001	1.001
0.01	0.5	0.998	0.981	0.936
	5.0	0.995	0.970	0.902

Case-Control Study (OR = 1.5, ρ = 1)		Values of ($OR_{EF/\bar{D}G}$, $OR_{EG/\bar{D}C}$)		
δ_{00}	$OR_{FG/D}$	(2,2)	(2,5)	(5,5)
0.01	0.5	1.003	0.985	0.939
	5.0	1.001	0.979	0.921
0.05	0.5	0.979	0.901	0.847
	5.0	0.966	0.868	0.763

TABLE 5.2

Average Relative Efficiency of Pooling and Stratification for Random Sampling in Case-Control Studies: No Confounding

$(OR = 1.5\ ,\ \delta_0 = 0.05\ ,\ OR_{FG/\bar{D}} = 2\ ,\ \rho = 1)$

$OR_{DG/\bar{E}F}$ \ $OR_{DF/\bar{E}G}$	0.0625	0.2	0.5	2	5	16
0.0625	2.25	1.84	1.59	1.48	1.55	1.87
0.2	1.79	1.35	1.20	1.16	1.26	1.58
0.5	1.61	1.19	1.06	1.05	1.16	1.50
2	1.47	1.15	1.05	1.07	1.21	1.58
5	1.51	1.24	1.16	1.20	1.36	1.80
16	1.83	1.55	1.49	1.59	1.80	2.29

TABLE 5.3

Average Relative Efficiency of Matching and Random Sampling: Residual Confounding Due to G, Follow-up Study ($OR_{EF/G}$=1)

(RR = 1.5 ; β_{00} = 0.0001 ; ρ = 1)

Matching on (a) F,G; and (b) F only

		Value of ($OR_{DF/\bar{E}G}$, $OR_{DG/\bar{E}F}$)					
		(2,2)		(2,5)		(5,5)	
$OR_{FG/E}$	$OR_{EG/F}$	(a)	(b)	(a)	(b)	(a)	(b)
0.5	0.0625	1.27	1.04	1.33	1.08	0.97	1.05
	0.2	1.02	1.02	1.04	1.04	0.87	1.03
	0.5	0.97	1.01	0.98	1.02	0.92	1.02
	2.0	1.08	0.99	1.07	0.98	1.15	0.99
	5.0	1.30	0.98	1.28	0.96	1.49	0.96
	16.0	1.83	0.98	1.78	0.94	2.18	0.93
5.0	0.0625	1.13	0.96	1.02	0.89	0.77	0.89
	0.2	0.95	0.96	0.90	0.92	0.79	0.91
	0.5	0.95	0.98	0.93	0.96	0.88	0.96
	2.0	1.11	1.03	1.13	1.05	1.18	1.06
	5.0	1.38	1.10	1.43	1.15	1.58	1.17
	16.0	2.04	1.21	2.19	1.32	2.52	1.38

TABLE 5.4

Average Relative Efficiency of Matching and Random Sampling:
Residual Confounding Due to G, Case-Control Study ($OR_{EF/\bar{D}G} = 1$)

($OR = 1.5$; $\delta_{00} = 0.05$; $\rho = 1$)

Matching on (a) F,G; and (b) F only

		$OR_{FG/D} = 0.5$				$OR_{FG/D} = 5.0$			
		$OR_{DF/\bar{E}G} = 2$		$OR_{DF/\bar{E}G} = 5$		$OR_{DF/\bar{E}G} = 2$		$OR_{DF/\bar{E}G} = 5$	
$OR_{EG/\bar{D}F}$	$OR_{DG/\bar{E}F}$	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
2.0	0.0625	1.28	1.03	1.35	1.07	—	—	—	—
	0.2	1.04	1.01	1.05	1.03	1.00	1.00	—	—
	0.5	0.98	1.00	0.98	1.00	0.98	1.00	0.99	1.02
	2.0	1.06	0.99	1.05	0.97	1.12	1.05	1.21	1.13
	5.0	1.26	0.98	1.23	0.95	1.38	1.12	1.52	1.21
	16.0	1.76	0.98	—	—	2.02	1.25	2.24	1.44
5.0	0.0625	1.05	1.02	1.10	1.04	—	—	—	—
	0.2	0.93	1.00	0.93	1.01	0.93	1.00	0.94	1.01
	0.5	0.94	0.99	0.93	0.99	0.96	1.01	0.99	1.05
	2.0	1.11	0.98	1.08	0.95	1.20	1.08	1.34	1.19
	5.0	1.41	0.98	1.35	0.93	1.55	1.16	1.76	1.30
	16.0	2.04	0.98	—	—	2.35	1.32	2.56	1.53

Dashes (-) indicate that no expected X^2 statistics could be generated

TABLE 5.5

Average Relative Efficiency of Matching and Random Sampling:
Joint Confounding, Follow-up Study

($RR = 1.5$; $B_{00} = 0.0001$; $OR_{FG/\bar{E}} = 1$; $OR_{DF/\bar{E}G} = 2$; $OR_{DG/\bar{E}F} = 2$)

Matching on (a) F,G; and (b) F only (in parentheses)

	OR(EG/F)					
OR(EF/G)	0.0625	0.2	0.5	2.0	5.0	16.0
0.0625	—	1.18 (1.18)	1.19 (1.24)	1.32 (1.21)	1.45 (1.11)	—
0.2	1.20 (0.99)	1.00 (1.00)	0.97 (1.00)	1.07 (0.99)	1.25 (0.97)	1.70 (0.93)
0.5	1.18 (0.97)	0.97 (0.97)	0.94 (0.97)	1.05 (0.96)	1.25 (0.96)	1.73 (0.95)
2.0	1.28 (1.07)	1.07 (1.08)	1.05 (1.09)	1.19 (1.09)	1.44 (1.09)	2.01 (1.09)
5.0	1.40 (1.22)	1.28 (1.29)	1.28 (1.32)	1.46 (1.33)	1.75 (1.32)	2.37 (1.27)
16.0	—	1.85 (1.84)	1.86 (1.93)	2.13 (1.94)	2.47 (1.87)	—

Dashes (—) indicate that no expected X^2 statistics could be generated.

TABLE 5.6

Average Relative Efficiency of Matching and Random Sampling:
Joint Confounding, Follow-up Study

($RR = 1.5$; $B_{00} = 0.0001$; $OR_{FG/\bar{E}} = 1$; $OR_{DF/\bar{E}G} = 5$; $OR_{DG/\bar{E}F} = 2$)

Matching on (a) F,G; and (b) F only (in parentheses)

	OR(EG/F)					
OR(EF/G)	0.0625	0.2	0.5	2.0	5.0	16.0
0.0625	—	0.90 (0.90)	0.89 (0.92)	0.96 (0.88)	1.06 (0.81)	—
0.2	1.04 (0.85	0.86 (0.86)	0.82 (0.85)	0.90 (0.83)	1.06 (0.82)	1.42 (0.78)
0.5	1.11 (0.91	0.91 (0.91)	0.88 (0.92)	0.98 (0.91)	1.16 (0.89)	1.59 (0.87)
2.0	1.35 (1.12)	1.13 (1.14)	1.11 (1.15)	1.27 (1.17)	1.54 (1.17)	2.17 (1.18)
5.0	1.50 (1.31)	1.43 (1.44)	1.46 (1.52)	1.69 (1.55)	2.02 (1.53)	2.75 (1.48)
16.0	—	2.27 (2.26)	2.28 (2.37)	2.68 (2.44)	3.23 (2.44)	—

Dashes (—) indicate that no expected X^2 statistics could be generated.

TABLE 5.7

Average Relative Efficiency of Matching and Random Sampling:
Joint Confounding, Follow-up Study

$(RR = 1.5\ ;\ B_{00} = 0.0001\ ;\ OR_{FG/\bar{E}} = 1\ ;\ OR_{DF/\bar{E}G} = 5\ ,\ OR_{DG/\bar{E}F} = 5)$

Matching on (a) F,G; and (b) F only (in parentheses)

	OR(EG/F)					
OR(EF/G)	0.0625	0.2	0.5	2.0	5.0	16.0
0.0625	—	0.80 (0.95)	0.85 (0.95)	0.99 (0.86)	1.15 (0.77)	—
0.2	0.80 (0.91)	0.75 (0.89)	0.78 (0.87)	0.95 (0.82)	1.16 (0.79)	1.57 (0.75)
0.5	0.83 (0.93)	0.78 (0.93)	0.83 (0.92)	1.04 (0.90)	1.30 (0.88)	1.80 (0.85)
2.0	0.96 (1.10)	0.95 (1.13)	1.04 (1.15)	1.36 (1.17)	1.79 (1.18)	2.60 (1.20)
5.0	1.08 (1.28)	1.19 (1.42)	1.35 (1.50)	1.83 (1.56)	2.40 (1.57)	3.51 (1.55)
16.0	—	1.82 (2.18)	2.09 (2.34)	2.80 (2.48)	3.74 (2.53)	—

Dashes (—) indicate that no expected X^2 statistics should be generated.

TABLE 5.8

Average Relative Efficiency of Matching and Random Sampling:
Joint Confounding, Follow-up Study

($RR = 1.5$; $B_{00} = 0.0001$; $OR_{FG/\bar{E}} = 5$, $OR_{DF/\bar{E}G} = 2$, $OR_{DG/EF} = 2$)

Matching on (a) F,G; and (b) F only (in parentheses)

	OR(EG/F)					
OR(EF/G)	0.0625	0.2	0.5	2.0	5.0	16.0
0.0625	—	1.32 (1.30)	1.15 (1.19)	1.09 (1.01)	—	—
0.2	1.32 (1.09)	1.00 1.00	0.94 (0.97)	1.00 (0.94)	1.15 (0.91)	—
0.5	1.14 (0.98)	0.94 (0.94)	0.92 (0.95)	1.03 (0.96)	1.22 (0.98)	1.80 (1.04)
2.0	—	1.00 (1.02)	1.03 (1.06)	1.26 (1.17)	1.61 (1.28)	2.44 (1.48)
5.0	—	—	1.25 (1.29)	1.65 (1.52)	2.12 (1.64)	3.14 (1.99)
16.0	—	—	—	2.50 (2.32)	3.34 (2.61)	—

Dashes (—) indicate that no expected X^2 statistics could be generated.

TABLE 5.9

Average Relative Efficiency of Matching and Random Sampling:
Joint Confounding, Follow-up Study

($RR = 1.5$; $B_{00} = 0.0001$; $OR_{FG/\overline{E}} = 5$; $OR_{DF/\overline{E}G} = 5$, $OR_{DG/\overline{E}F} = 2$)

Matching on (a) F,G; and (b) F only (in parentheses)

	OR(EG/F)					
OR(EF/G	0.0625	0.2	0.5	2.0	5.0	16.0
0.0625	—	0.94 (0.94)	0.84 (0.87)	0.83 (0.77)	—	—
0.2	1.01 (0.85)	0.82 (0.83)	0.79 (0.82)	0.88 (0.83)	1.02 (0.83)	—
0.5	0.98 (0.86)	0.85 (0.86)	0.85 (0.88)	0.98 (0.92)	1.18 (0.97)	1.81 (1.07)
2.0	—	1.01 (1.03)	1.07 (1.10)	1.35 (1.26)	1.80 (1.44)	2.79 (1.73)
5.0	—	—	1.38 (1.43)	1.92 (1.78)	2.51 (1.95)	3.87 (2.51)
16.0	—	—	—	3.07 (2.85)	4.33 (3.39)	—

Dashes (—) indicate that no expected X^2 statistics could be generated.

TABLE 5.10

Average Relative Efficiency of Matching and Random Sampling:
Joint Confounding, Follow-up Study

($RR = 1.5$; $B_{00} = 0.0001$; $OR_{FG/\bar{F}} = 5$; $OR_{DF/\bar{E}G} = 5$, $OR_{DG/\bar{E}F} = 5$)

Matching on (a) F,G; and (b) F only (in parentheses)

	OR(EF/F)					
OR(EF/G)	0.0625	0.2	0.5	2.0	5.0	16.0
0.0625	—	0.76 (0.89)	0.75 (0.82	0.81 (0.72)	—	—
0.2	0.76 (0.85)	0.70 (0.82)	0.74 (0.80)	0.88 (0.80)	1.07 (0.80)	—
0.5	0.74 (0.85)	0.73 (0.85)	0.80 (0.87)	1.01 (0.91)	1.26 (0.97)	2.04 (1.08)
2.0	—	0.89 (1.03)	1.02 (1.11)	1.44 (1.29)	2.03 (1.51)	3.24 (1.85)
5.0	—	—	1.34 (1.45)	2.11 (1.86)	2.99 (2.10)	4.47 (2.74)
16.0	—	—	—	3.35 (3.00)	5.15 (3.74)	—

Dashes (—) indicate that no expected X^2 statistics could be generated.

TABLE 5.11

Average Relative Efficiency of Matching and Random Sampling: Joint Confounding, Case-Control Study

($OR = 1.5$; $\delta_{00} = 0.05$; $OR_{FG/\bar{D}} = 1$; $OR_{EF/\bar{D}G} = 2$, $OR_{EG/\bar{D}F} = 2$)

Matching on (a) F,G; and (b) F only (in parenthesss)

$OR(DF/\bar{E}G)$	$OR(DG/\bar{E}F)$ 0.0625	0.2	0.5	2.0	5.0	16.0
0.0625	—	1.28 (1.23)	1.27 (1.30)	1.40 (1.30)	1.54 (1.21)	—
0.2	1.30 (1.01)	1.07 (1.03)	1.02 (1.04)	1.11 (1.03)	1.28 (1.02)	1.72 (0.98)
0.5	1.27 (0.98)	1.02 (0.98)	0.97 (0.98)	1.05 (0.98)	1.24 (0.98)	1.66 (0.97)
2.0	1.38 (1.06)	1.11 (1.07)	1.06 (1.07)	1.15 (1.08)	1.36 (1.08)	1.88 (1.07)
5.0	1.51 (1.20)	1.31 (1.26)	1.26 (1.23)	1.38 (1.28)	1.61 (1.26)	2.15 (1.23)
16.0	—	1.84 (1.75)	1.79 (1.82)	1.97 (1.82)	2.23 (1.74)	—

Dashes (—) indicate that no expected X^2 statistics could be generated.

TABLE 5.12

Average Relative Efficiency of Matching and Random Sampling:
Joint Confounding, Case-Control Study

$(OR = 1.5 \; ; \quad \delta_{00} = 0.05 \; ; \quad OR_{FG/\bar{D}} = 1 \; , \quad OR_{EF/\bar{D}G} = 5 \; , \quad OR_{EG/\bar{D}F} = 2)$

Matching on (a) F,G; and (b) F only (in parentheses)

	$OR(DG/\bar{E}F)$					
$OR(DF/\bar{E}F)$	0.0625	0.2	0.5	2.0	5.0	16.0
0.0625	—	1.12 (1.06)	1.11 (1.12)	1.22 (1.14)	1.33 (1.06)	—
0.2	1.23 (0.91)	1.01 (0.95)	0.95 (0.96)	1.01 (0.96)	1.16 (0.95)	1.54 (0.91)
0.5	1.27 (0.94)	1.03 (0.96)	0.96 (0.96)	1.01 (0.96)	1.17 (0.95	1.56 (0.94)
2.0	1.50 (1.11)	1.20 (1.11)	1.12 (1.12)	1.18 (1.12)	1.37 (1.12)	1.87 (1.13)
5.0	1.75 (1.28)	1.47 (1.36)	1.40 (1.40)	1.48 (1.40)	1.66 (1.37)	2.25 (1.37)
16.0	—	2.24 (2.03)	2.09 (2.08)	2.21 (2.09)	2.46 (1.99)	—

Dashes (—) indicate that no expected X^2 statistics could be generated.

TABLE 5.13

Average Relative Efficiency of Matching and Random Sampling:
Joint Confounding, Case-Control Study

($OR = 1.5$; $\delta_{00} = 0.05$; $OR_{FG/\bar{D}} = 1$, $OR_{EF/\bar{D}G} = 5$, $OR_{EG/\bar{D}F} = 5$

Matching on (a) F,G; and (b) F only (in parentheses)

$OR(DF/\bar{E}G)$	$OR(DG/\bar{E}F)$ 0.0625	0.2	0.5	2.0	5.0	16.0
0.0625	—	1.14 (1.16)	1.19 (1.24)	1.47 (1.32)	1.69 (1.24)	—
0.2	1.16 (0.96)	1.01 (1.00)	1.00 (1.03)	1.13 (1.04)	1.33 (1.04)	1.86 (1.01)
0.5	1.18 (0.96)	1.01 (0.98)	0.97 (0.98)	1.06 (0.99)	1.27 (0.99)	1.74 (0.99)
2.0	1.41 (1.08)	1.13 (1.08)	1.07 (1.08)	1.16 (1.08)	1.36 (1.08)	1.88 (1.07)
5.0	1.65 (1.23)	1.37 (1.28)	1.29 (1.30)	1.38 (1.29)	1.57 (1.26)	—
16.0	—	2.03 (1.86)	1.87 (1.87)	1.94 (1.82)	—	—

Dashes (—) indicate that no expected X^2 statistics could be generated.

TABLE 5.14

Average Relative Efficiency of Matching and Random Sampling:
Joint Confounding, Case-Control Study

$(OR = 1.5\ ;\ \delta_{00} = 0.05\ ;\ OR_{FG/\bar{D}} = 5\ ,\ OR_{EF/\bar{D}G} = 2\ ,\ OR_{EG/\bar{D}F} = 2)$

Matching on (a) F,G; and (b) F only (in parentheses)

	$OR(DG/\bar{E}F)$					
$OR(DF/\bar{E}G)$	0.0625	0.2	0.5	2.0	5.0	16.0
0.0625	—	1.47 (1.41)	1.28 (1.30)	1.20 (1.11)	—	—
0.2	—	1.08 (1.04)	1.01 (1.03)	1.04 (0.99)	1.17 (0.95)	—
0.5	1.22 (1.00)	1.00 (0.97)	0.96 (0.97)	1.03 (0.98)	1.19 (0.99)	1.73 (1.04)
2.	—	1.04 (1.01)	1.04 (1.05)	1.21 (1.14)	1.49 (1.22)	2.16 (1.37)
5.0	—	—	1.23 (1.24)	1.54 (1.44)	1.89 (1.54)	2.71 (1.80)
16.0	—	—	—	2.29 (2.15)	2.96 (2.38)	—

Dashes (—) indicate that no expected X^2 statistics could be generated.

TABLE 5.15

Average Relative Efficiency of Matching and Random Sampling:
Joint Confounding, Case-Control Study

($OR = 1.5$; $\delta_{00} = 0.05$; $OR_{FG/\bar{D}} = 5$; $OR_{EF/\bar{D}\bar{G}} = 5$, $OR_{EG/\bar{D}\bar{F}} = 2$)

Matching on (a) F,G; and (b) F only (in parentheses)

	$OR(DG/\bar{E}F)$					
$OR(DF/\bar{E}G)$	0.0625	0.2	0.5	2.0	5.0	16.0
0.0625	—	1.27 (1.21)	1.12 (1.13)	1.05 (0.99)	—	—
0.2	—	1.01 (0.95)	0.94 (0.95)	0.98 (0.93)	1.10 (0.92)	—
0.5	1.21 (0.94)	0.98 (0.93)	0.94 (0.94)	1.01 (0.96)	1.17 (0.99)	1.70 (1.06)
2.0	—	1.09 (1.02)	1.08 (1.08)	1.24 (1.19)	1.53 (1.30)	2.24 (1.48)
5.0	—	—	1.33 (1.33)	1.66 (1.58)	1.98 (1.67)	2.90 (2.00)
16.0	—	–	—	2.57 (2.44)	—	—

Dashes (—) indicate that no expected X^2 statistics could be generated.

TABLE 5.16

Average Relative Efficiency of Matching and Random Sampling:
Joint Confounding, Case-Control Study

$(OR = 1.5 \; ; \; \delta_{00} = 0.05 \; ; \; OR_{FG/\overline{D}} = 5 \; ; \; OF_{EF/\overline{DG}} = 5 \; , \; OR_{EG/\overline{DF}} = 5)$

Matching on (a) F,G; and (B) F only (in parentheses)

$OR(DF/\overline{E}G)$	$OR(DG/\overline{E}F)$ 0.0625	0.2	0.5	2.0	5.0	16.0
0.0625	1.43 (1.30)	1.26 (1.29)	1.20 (1.25)	1.24 (1.11)	—	—
0.2	—	1.02 (1.00)	0.99 (1.02)	1.07 (0.99)	1.24 (0.96)	—
0.5	1.17 (0.97)	0.99 (0.96)	0.95 (0.96)	1.05 (0.98)	1.25 (1.00)	1.80 (1.05)
2.0	—	1.06 (1.02)	1.05 (1.07)	1.22 (1.15)	1.51 (1.23)	2.15 (1.38)
5.0	—	1.17 (1.13)	1.27 (1.28)	1.59 (1.49)	1.84 (1.55)	2.71 (1.79)
16.0	—	—	—	2.41 (2.27)	—	—

Dashes (—) indicate that no expected X^2 statistics could be generated.

TABLE 5.17

Average Relative Efficiency of Matching and Random Sampling:
Joint Confounding, Case-Control Study

Matching on F and G Incorporating Loss of Sample Size (Method B)

($OR = 1.5$; $\delta_{00} = 0.05$; $OR_{FG/\bar{D}} = 1$; $OR_{EF/\bar{D}G} = 2$; $OR_{EG/\bar{D}F} = 2$)

$OR(DF/\bar{E}G)$	$OR(DG/\bar{E}F)$ 0.0625	0.2	0.5	2.0	5.0	16.0
0.0625	—	0.22	0.38	0.40	0.37	—
0.2	0.24	0.36	0.46	0.44	0.35	0.41
0.5	0.41	0.48	0.63	0.51	0.49	0.47
2.0	0.38	0.51	0.66	0.66	0.54	0.46
5.0	0.33	0.41	0.53	0.56	0.44	—
16.0	—	0.36	0.44	0.46	—	—

Dashes (—) indicate that no expected X^2 statistics could be generated.

CHAPTER 6

SUMMARY AND SUGGESTIONS FOR FUTURE RESEARCH

6.1 Summary

The primary intent of this work has been to evaluate the effectiveness of (frequency-) matching in providing a more efficient analysis of the association between a disease variable and exposure agent with respect to random sampling. Particular consideration was given to the characterization of the relationship betweeen the relative efficiency of matching vs. random sampling and the degree and direction of the underlying confounding. Relative efficiency was considered in the context of a single dichotomous matching variable and two dichotomous matching variables. The results of the evaluation show that, in general, there is an expected increase in efficiency from matching on a confounding variable. The increase is larger for the follow-up study but more likely for the case-control study. No loss in efficiency results from matching on non-confounders unless they are highly related to "negative" confounders.

Evaluation of the relative efficiency was based on analytical and numerical studies of ratios of Mantel-Haenszel X^2 statistics constructed from "expected" cell frequencies under the two designs. These ratios, when evaluated near the null values of the measures of effect, are shown to be equivalent to the asymptotic relative efficiency, or Pitman efficiency (Gibbons, 1971).

In order to incorporate loss of sample size from the matching into the evaluation of the relative efficiency, algorithms were developed to quantify the loss of sample size as functions of the joint distributions of the matching variables. In that regard, the loss of sample size is seen to vary with the strength of confounding. Regardless of the formulation of the loss in sample size, matching is shown to lead to a consistent, substantial loss in efficiency relative to random sampling.

In the context of categorical matching variables, the method of pair-matching is compared to frequency-matching with respect to validity and efficiency. It is demonstrated that the pair-matched odds ratio is a function of the pairing, and due to the arbitrary nature of the pairing, can be grossly inaccurate with respect to the true value of the odds ratio. In addition, pair-matching is seen to result in a less efficient analysis. It follows that pairing is rejected as a method of matching in this context.

In addition to efficiency considerations, conditions for no confounding are developed for each study type assuming two potential confounding variables. The concepts of "joint" and "residual" confounding are discussed in detail. Conditions are developed through comparisons of crude and adjusted measures of effect as well as a regression formulation. The concept of confounding as it pertains to two potential confounders is seen to grow more elusive. In that regard the proper identification of a (non-) confounder is drawn. Attention is also given to the effectiveness of matching as a design technique to control confounding, and to the manifestation of apparent effect modification as a form of confounding, which is demonstrated for the case of two confounding variables.

6.2 Suggestions for Future Research

Extensions and further developments of the present research which might be warranted include the following:

(a) Further investigation of the relationships between the underlying confounding and the relative efficiency.

A key result of this dissertation is the identification of a relationship between the underlying confounding and the efficiency of matching. It would be of interest to continue the investigation concentrating on the relationship between the size of the confounding bias and the relative efficiency. By focusing on the size of the bias rather than associations between potential confounders and disease and exposure, some of the restrictive assumptions that were required for the present work can be dropped, in particular, the "no interaction" restrictions.

(b) Strategies for choosing the "best" set of confounders to control.

Kleinbaum and Kupper (1980) raise the possibility of a circumstance in which a confounding bias can be removed by adjusting for different sets of confounders: that is, unique sets of confounders may be "responsible" for the same bias. Such a circumstance was presented in Chapter 4: one of two extraneous variables, themselves correlated, is independent of disease while the other is independent of exposure. Controlling for either removes confounding; controlling for neither leaves a confounding bias. With the possibility that the control of confounding can by accomplished by adjusting for different sets of confounders, it would be useful to develop a strategy to choose the "best" set (or, minimal number) of confounders for which control is to be accomplished. Such a strategy would necessarily involve efficiency

considerations.

(c) Extension of the relative efficiency evaluations to continuous variables and continuous-variable matching methods.

The efficiency of various continuous-variable matching methods, such a nearest-neighbor and caliper matching, has been studied relative to analytical methods (e.g., analysis of covariance) by a number of authors including Billewicz (1965), Rubin (1973), and Raynor and Kupper (1977). None of the techniques has been applied to more than a single matching variable and compared with respect to an analytical method of control. Rubin (1976) and Miettinen (1976) have advocated various multivariable methods which reduce a set of matching variables via a linear function to a single "score" which is then used as the matching variable. Raynor and Kupper (1977) emphasize that there remains a need for a more refined understanding of the performance of continuous-variable matching techniques, both in terms of their effectiveness to control confounding as well as their efficiency vis-a-vis methods such as analysis of covariance.

BIBLIOGRAPHY

Billewicz, W.Z. (1965), "The Efficiency of Matched Samples: An Empirical Investigation." Biometrics, 21, 623-643.

Bross, I.D.J. (1966), "Spinous Effects from an Extraneous Variable." Journal of Chronic Disease, 19, 637-647.

Bross, I.D.J. (1967), "Pertinency of an Extraneous Variable." Journal of Chronic Disease, 20, 487-495.

Bross, I.D.J. (1969), "How Case-for-Case Matching can Improve Design Efficiency." American Journal of Epidemiology, 89, 359-363.

Chase, G.R. (1968), "On the Efficiency of Matched Pairs in Bernoulli Trails." Biometrika, 55, 365-369.

Cochran, W.G. (1953), "Matching in Analytical Studies." American Journal of Public Health, 43, 684-691.

Cochran, W.G. (1965), "The Planning of Observational Studies of Human Populations." Journal of the Royal Statistical Society, Series A, 128, 234-255.

Cornfield, J. (1956), "A Statistical Problem Arising from Retrospective Studies." Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, 135-148.

Cornfield, J.; Haenszel, W.; Hammond, E.C.; Lilienfeld, A.M.; Shimkin, M.B.; and Wynder, E.L. (1959), "Smoking and Lung Cancer: Recent Evidence and Discussion of Some Questions." Journal of National Cancer Institute, 22, 1973-203.

Fisher, L. and Patil, K. (1974), "Matching and Unrelatedness." American Journal of Epidemiology, 100, 347-349.

Gibbons, J.D. (1971), Non-Parametric Statistical Inference, McGraw-Hill, 276-281.

Hamilton, M.A. (1979), "Choosing the Parameter for a 2 x 2 Table or a 2 x 2 x 2 Table Analysis." American Journal of Epidemiology, 109, 362-375.

Kleinbaum, D.G. and Kupper, L.L. (1980), "Confounding in the Presence of Several Extraneous Factors." Unpublished manuscript.

Kraus, A.S. (1958), "The Use of Family Members as Controls in the Study of the Possible Etiologic Factors of a Disease." Ph.D. Dissertation, Department of Biostatistics, School of Public Health, University of Pittsburgh.

Kupper, L.L.; Karon, J.M.; Kleinbaum, D.G.; Morganstern, H.; and Lewis, D.K. (1980). "Matching in Epidemiologic Studies: Validity and Efficiency Considerations." To appear in Biometrics.

Mantel, N. and Haenszel, W. (1959), "Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease." Journal of National Cancer Institute, 22, 719-748.

Mathen, K.K. (1963), "Matching in Comparative Studies in Public Health." Indian Journal of Public Health, 7, 161-169.

McKinlay, S.M. (1975a), "A Note on the Chi-Square Test for Pair-Matched Samples." Biometrics, 31, 731-735.

McKinlay, S.M. (1975b), "The Design and Analysis of the Observational Study: A Review." Journal of the American Statistical Association, 70, 503-520.

McKinlay, S.M. (1975c), "The Effect of Bias on Estimators of Relative Risk for Pair-Matched and Stratified Samples." Journal of the American Statistical Association, 70, 859-864.

McKinlay, S.M. (1977), "Pair-Matching — A Reappraisal of a Popular Technique." Biometrics, 33, 725-735.

Miettinen, O. (1968), "The Matched Pairs Design in the Case of All-or-None Response." Biometrics, 24, 339-352.

Miettinen, O. (1969), "Individual Matching with Multiple Controls in the Case of All-or-None Response." Biometrics, 25, 339-355.

Miettinen, O. (1970), "Matching and Design Efficiency in Retrospective Studies." American Journal of Epidemiology, 91, 111-118.

Miettinen, O. (1972a), "Components of the Crude Risk Ratio." American Journal of Epidemiology, 96, 168-172.

Miettinen, O. (1972b), "Standardization of Risk Ratios." American Journal of Epidemiology, 96, 383-388.

Miettinen, O. (1974), "Confounding and Effect Modification." American Journal of Epidemiology, 100, 350-353.

Miettinen, O. (1975), "Principles of Epidemiologic Research. Unpublished manuscript.

Miettinen, O. (1976), "Estimability and Estimation in Case-Referent Studies." American Journal of Epidemiology, 103, 226-235.

Miettinen, O. (1976), "Stratification by a Multivariate Confounder Score." American Journal of Epidemiology, 104, 609-620.

Miettinen, O. and Cook (1980), Unpublished manuscript.

Pike, M.C. and Morrow, R.H. (1970), "Statistical Analysis of Patient Control Studies in Epidemiology." British Journal of Preventive and Social Medicine, 24, 42-44.

Raynor, W.J. and Kupper, L.L. (1977), "Matching on a Continuous Variable in Case-Control Studies." Department of Biostatistics, University of North Carolina, Institute Statistics Mimeo Series No. 1127.

Rubin (1976), "Multivariate Matching Methods that are Equal Percent Bias Reducing." Biometrics, 32, 109-132.

Samuels, M.L. (1979), "How Case Matching can Reduce Design Efficiency in Retrospective Studies." Department of Statistics, Purdue University Mimeograph Series #79-30.

Samuels, M.L. (1980), "The Relative Efficiency of Matching versus Post-Stratification for Bias Control in Epidemiological Studies." Department of Statistics, Purdue University Mimeograph Series #80-15.

Schlesselman, J.J. (1978). "Assessing Effects of Confounding Variables." American Journal of Epidemiology, 108, 3-8.

Siegel, D.G. and Greenhouse, S.W. (1975), "Validity in Estimating Relative Risk in Case-Control Studies." Journal of Chronic Disease, 27, 216-225.

Stuart, A. (1957), "The Comparison of Frequencies in Matched Samples." British Journal of Statistical Psychology, 10, 29-32.

Wold, H. (1956), "Causal Inference from Observational Data." Journal of the Royal Statistical Society, Series A, 119, 28-61.

Worcester, J. (1964), "Matched Samples in Epidemiologic Studies." Biometrics, 20, 840-848.

APPENDIX 1

The Relative Efficiency of X^2_{McN} and X^2_{MH} When the Sample Size and the Measure of Effect are Small

In Chapter 2 pair-matching is shown to be less efficient than an analysis which does not utilize the pairing. The expected Mantel-Haenszel statistic, X^2_{MH}, is shown to be larger then the expected McNemar statistic, X^2_{McN}, when the sample size is assumed large (2.3.4). However, when the sample size is not large, an evaluation of the relative efficiency cannot involve the substitution in (2.3.2) of $2n\theta_k-1$ with $2n\theta_k$. The impact of this substitution on the relative efficiency is now studied for small sample sizes and measures of effect which approach the null value.

Without this substitution the relative efficiency (2.3.5) is expressed as

$$RE = \frac{n\sum_k \theta_k^2(\alpha_k+\beta_k)(2-(\alpha_k+\beta_k))/(2n\theta_k-1)}{\sum_k \theta_k(\alpha_k+\beta_k-2\alpha_k\beta_k)} . \qquad (1)$$

For pair-matching to be more efficient, $RE < 1$, or

$$\sum_k \theta_k(\alpha_k+\beta_k-2\alpha_k\beta_k) > n\sum_k \theta_k^2(\alpha_k+\beta_k)(2-\alpha_k-\beta_k)/(2n\theta_k-1) ,$$

$$> 2n\sum_k \frac{\theta_k^2(\alpha_k+\beta_k)}{2n\theta_k-1} - n\sum_k \frac{\theta_k^2(\alpha_k+\beta_k)^2}{2n\theta_k-1} ,$$

or equivalently,

$$\sum_k \theta_k(\alpha_k+\beta_k)\left(1 - \frac{2n\theta_k}{2n\theta_k-1}\right) > \sum \frac{\theta_k}{2n\theta_k-1}\left\{2\alpha_k\beta_k(2n\theta_k-1) - n\theta_k(\alpha_k+\beta_k)^2\right\} ,$$

or

$$- \sum_k \frac{\theta_k(\alpha_k+\beta_k)}{2n\theta_k-1} > \sum_k \frac{\theta_k}{2n\theta_k-1}\left\{-n\theta_k(\alpha_k^2-2\alpha_k\beta_k+\beta_k^2) - 2\alpha_k\beta_k\right\}$$

$$> - \sum_k \frac{\theta_k}{2n\theta_k-1}\left\{n\theta_k(\alpha_k+\beta_k)^2 + 2\alpha_k\beta_k\right\} ,$$

which can also be written as

$$\sum_k \frac{\theta_k}{2n\theta_k-1}\left\{n\theta_k(\alpha_k-\beta_k)^2 - (\alpha_k+\beta_k-2\alpha_k\beta_k)\right\} > 0 . \qquad (2)$$

For the follow-up study, assuming uniformity of the risk ratio, $\alpha_k = RR\cdot\beta_k$, for $\forall k$, (2) can be expressed as

$$\sum_k \frac{\beta_k\theta_k}{2n\theta_k-1}\left\{n\theta_k\beta_k(RR-1)^2 - (RR+1-2\beta_k RR)\right\} > 0 . \qquad (3)$$

Consider the term in brackets. This term must be positive in a large number of strata for (3) to hold. For the k^{th} term to be positive,

$$n\theta_k\beta_k(RR-1)^2 > RR+1 - 2\beta_k RR$$

or

$$\beta_k > (RR+1)/(n\theta_k(RR-1)^2 + 2RR) . \qquad (4)$$

A table of minimum values of β_k which will yield a positive term in (3) are given below for various values of RR, n, and θ_k.

TABLE 1

	RR = 2.0				RR = 1.5			
n	θ_k: 0.05	0.10	0.20	0.30	θ_k: 0.05	0.10	0.20	0.30
100	0.333	0.214	0.125	0.088	0.588	0.455	0.313	0.238
200	0.214	0.125	0.068	0.047	0.455	0.313	0.192	0.139
500	0.103	0.055	0.029	0.019	0.270	0.161	0.089	0.061
1000	0.055	0.029	0.015	0.010	0.161	0.089	0.047	0.032

Clearly, these values for β_k are unrealistically large for follow-up studies involving rare diseases. Only when the sample size is large do the minimum values of β_k approach levels which might be expected in practice. Therefore, the large-sample assumption accompanying the relative efficiency results in Chapter 2 as regards the follow-up study can be expected to hold in practice for small samples.

For the case-control study the properties of the relative efficiency under small measures of effect and sample size can be evaluated in much the same manner. Consider (2) expressed in terms of the case-control parameters as follows.

$$\sum_k \frac{v_k}{2nv_k-1}\left\{nv_k\delta_k(\varepsilon_k-\delta_k)^2-(\varepsilon_k+\delta_k-2\varepsilon_k\delta_k)\right\} > 0 \ . \qquad (5)$$

Substituting $\varepsilon_k = \delta_k \cdot OR/(1+\delta_k(OR-1))$, (5) is expressed as

$$\sum_k \frac{v_k\delta_k}{2nv_k-1}\left\{nv_k\delta_k\left(\frac{OR}{1+\delta_k(OR-1)}-1\right)^2-\left(\frac{OR}{1+\delta_k(OR-1)}+1-\frac{2\ OR\ \delta_k}{1+\delta_k(OR-1)}\right)\right\} > 0$$

or,

$$\sum_k \frac{v_k\delta_k}{2nv_k-1}\left\{nv_k\delta_k\ \frac{(OR-1)^2(1-\delta_k)^2}{(1+\delta_k(OR-1))^2}-\frac{(OR+1)(1-\delta_k)}{(1+\delta_k(OR-1))}\right\} > 0$$

or,

$$\sum_k \frac{v_k \delta_k (1-\delta_k)}{(2nv_k - 1)(1+\delta_k(OR-1))} \left\{ \frac{nv_k \delta_k (1-\delta_k)(OR-1)^2}{1+\delta_k(OR-1)} - (OR+1) \right\} > 0 \ . \qquad (6)$$

Consider again the term in brackets. For (6) to hold this term must be positive in a majority of strata. If this is true for the k^{th} strata, then

$$\frac{nv_k \delta_k (1-\delta_k)(OR-1)^2}{1+\delta_k(OR-1)} > OR + 1$$

or,

$$\frac{\delta_k (1-\delta_k)}{(OR-1)^{-1} + \delta_k} > \frac{OR+1}{(OR-1)nv_k} = c \ .$$

Equivalently,

$$\delta_k - \delta_k^2 > c\delta_k + c(OR-1)^{-1}$$

$$-\delta_k^2 + (1-c)\delta_k - c(OR-1)^{-1} > 0 \ . \qquad (7)$$

Let $f(\delta_k) = -\delta_k^2 + (1-c)\delta_k - c(OR-1)^{-1}$. The function $f(\cdot)$ is a quadratic function of δ_k, and is negative for extreme values of δ_k (near 0 or 1). For $f(\delta_k) > 0$

$$\delta_k > \frac{1-c}{2} - \frac{\sqrt{A}}{2}$$

and,

$$\delta_k < \frac{1-c}{2} + \frac{\sqrt{A}}{2} \ , \qquad (8)$$

where

$$A = 1-2c+c^2-4c/(OR-1) \ .$$

The term A must be positive for there to exist values of δ_k which will allow $f(\cdot) > 0$. Now

$$A = 1 - 2c + c^2 - 4c/(OR-1)$$

$$= 1 - \frac{2(OR+1)}{(OR-1)nv_k} + \frac{(OR+1)^2}{(OR-1)^2 n^2 v_k^2} - \frac{4(OR+1)}{(OR-1)^2 nv_k} .$$

Consider values of A for $OR = 2.0$ and $OR = 1.5$.

For $OR = 2.0$,

$$A = 1 - 6/nv_k + 9/n^2v_k^2 - 12/nv_k$$

$$= 1 - 18/nv_k + 9/n^2v_k^2 .$$

$A > 0$ if $(nv_k)^2 - 18nv_k + 9 > 0$ or,

$$nv_k > \frac{18 + \sqrt{324-36}}{2} \simeq 17.5 .$$

For $OR = 1.5$, $A = 1 - 10/nv_k + 25/n^2v_k^2 - 40/nv_k = 1 - 50/nv_k + 25/n^2v_k^2$.

$A > 0$ if $(nv_k)^2 - 50nv_k + 25 > 0$ or,

$$nv_k > \frac{50 + \sqrt{2500-100}}{2} \simeq 49.5 .$$

Hence, for the existence of a δ_k which satisfies (7), the number of cases in each strata must exceed 18 for $OR = 2.0$ and 50 for $OR = 1.5$. And, for a range of δ_k to satisfy (7), the stratum-specific number of cases must be even larger. Note that the midpoint of such a range is given by (8),

$$\delta_k = \frac{1-c}{2} = \tfrac{1}{2}\left(1 - \frac{OR+1}{(OR-1)nv_k}\right)$$

$$= \begin{cases} 0.5 - 3/2nv_k , & OR = 2.0 \\ 0.5 - 5/2nv_k , & OR = 1.5 , \end{cases}$$

which varies between 0.40 and 0.45. These would be rather high

probabilities of exposure among controls for most practical situations. The range of δ_k considered in Chapter 2 had an upward limit of 0.36.

All things considered, there is no apparent advantage in terms of efficiency to pair-matching over frequency-matching in case-control studies when both the sample size and the odds ratio are small.

APPENDIX 2

In this appendix to Chapter 3 the index of relative efficiency, RE (3.3.5), is proven to be equivalent asymptotically to the Pitman efficiency (Noether, 1950).

Consider Z, where $Z = (X^2)^{\frac{1}{2}}$, the root of the observed M-H X_1^2 statistic (3.3.1). Let Z be expressed by dividing both numerator and denominator by $N^{\frac{1}{2}}$. Therefore,

$$Z = \frac{N^{-\frac{1}{2}} \left(\sum_i \frac{a_i b_i - c_i d_i}{N_i} \right)}{\left(\sum \frac{N_{1i} N_{2i} M_{1i} M_{2i}}{N N_i^2 (N_i - 1)} \right)^{\frac{1}{2}}} .$$

Assuming that the marginal frequencies of each stratum $(N_{1i}, N_{2i}, M_{1i}, M_{2i})$ are fixed, then

$$N^{-\frac{1}{2}} \sum_i \frac{a_i d_i - b_i c_i}{N_i} = N^{-\frac{1}{2}} \left(\sum_i \left(a_i - \frac{N_{1i} M_{1i}}{N_i} \right) \right)$$

$$= N^{\frac{1}{2}} \left(\bar{a} - \sum_i \frac{N_{1i} M_{1i}}{N N_i} \right)$$

where $$\bar{a} = \sum_i a_i / N .$$

Also,

$$\sum_i \frac{N_{1i} M_{1i}}{N N_i} = N^{-1} \sum_i E(a_i) = E(\bar{a}) = \mu ,$$

and

$$\left(\sum_i \frac{N_{1i} N_{2i} M_{1i} M_{2i}}{N N_i^2 (N_i - 1)} \right)^{\frac{1}{2}} = \left(\frac{1}{n} \sum_i \mathrm{Var}(a_i) \right)^{\frac{1}{2}} = \left(\frac{1}{N} \mathrm{Var}(\bar{a}) \right)^{\frac{1}{2}} = \sigma .$$

Therefore, $Z = N^{\frac{1}{2}} (\bar{a} - \mu) / \sigma$.

Conditional on $(N_{1i},N_{2i},M_{1i},M_{2i})$, the a_i are independent. Therefore, by application of the Central Limit Theorem, $Z \to N(0,1)$ as $N \to \infty$.

Let Z and Z^* represent the test statistics for random sampling and matching, respectively. Both Z and Z^* are asymptotically standard normal, where

$$Z^* = N^{\frac{1}{2}}(\bar{a} - \mu^*)/\sigma^* .$$

Assume without loss of generality that the context of this proof is the follow-up study. Under H_0,

$$E(Z) = E(Z^*) = 0 .$$

Also,

$$\mu = \mu^* = \sum_i \alpha_i \theta_i .$$

Under H_A

$$E(Z) = \frac{N^{\frac{1}{2}}}{\sigma} (E(\bar{a}|H_A) - \mu)$$

$$= \frac{N^{\frac{1}{2}}}{\alpha} \left[\sum_i (\alpha_i\theta_i+\beta_i\phi_i)\theta_i/(\theta_i+\phi_i) - \sum_i \alpha_i\theta_i \right]$$

$$= \frac{N^{\frac{1}{2}}}{\sigma} \left[- \sum_i \frac{\theta_i\phi_i(\alpha_i-\beta_i)}{\theta_i+\phi_i} \right] ,$$

$$E(Z^*) = \frac{N^{\frac{1}{2}}}{\sigma^*} \left[\sum_i \frac{\theta_i^2(\alpha_i-\beta_i)}{2\theta_i} - \sum_i \alpha_i\theta_i \right]$$

$$= \frac{N^{\frac{1}{2}}}{\sigma^*} \left[- \sum_i \frac{\theta_i(\alpha_i-\beta_i)}{2} \right] .$$

Now, let

$$E(Z) = \frac{1}{\sigma} \left[- \sum_i \theta_i\phi_i\beta_i\delta_i/(\theta_i+\phi_i) \right] ,$$

and

$$E(Z^*) = \frac{1}{*}\left[-\sum_i \theta_i \beta_i \delta_i / 2\right] ,$$

where

$$N^{\frac{1}{2}}\delta_i = \frac{\alpha_i}{\beta_i} - 1 = RR_i - 1 .$$

Let $\delta_i = c_i\delta$, where c_i is fixed. Then, replace c_i with

$$\bar{c} + \gamma_i , \quad \text{where} \quad \sum_i \gamma_i = 0 .$$

Since $\delta_i > 0$, $\forall i$, then $c_i > 0$ and $\bar{c} > 0$. Therefore,

$$\delta_i = \delta\bar{c} + \delta\gamma_i ,$$

and

$$E(Z) = \frac{1}{\sigma}\left[-\sum_i \frac{\theta_i\phi_i\beta_i}{\theta_i+\phi_i}(\delta\bar{c} + \delta\gamma_i)\right]$$

$$= \frac{1}{\sigma}\left[-\delta\bar{c}\sum_i \frac{\theta_i\phi_i\beta_i}{(\theta_i+\phi_i)} - \delta\sum_i \frac{\theta_i\phi_i\beta_i\gamma_i}{\theta_i+\phi_i}\right] .$$

Note that δ can be considered an "average" alternative to the null, $\delta = 0$. Assuming there is only minor inter-stratum variability in δ_i, then the γ_i can be considered "small." And, because $\sum_i \gamma_i = 0$, the second term in brackets above can be considered negligible, and

$$E(Z) \approx -\frac{\delta\bar{c}}{\sigma}\sum_i \frac{\theta_i\phi_i\beta_i}{(\theta_i+\phi_i)} = u(\delta) .$$

Similarly,

$$E(Z^*) \approx -\frac{\delta\bar{c}}{\sigma^*}\sum_i \theta_i\beta_i/2 = u^*(\delta) .$$

Hence, Z and Z^* are implicitly expressed as

$$Z = \frac{t-u(\delta)}{v(\delta)} \quad \text{and} \quad Z^* = \frac{t-u^*(\delta)}{v(\delta)} ,$$

where

$$v^2(\delta) = \text{Var}(Z)/N = 1/N$$

$$v^{*2}(\delta) = \text{Var}(Z^*)/N = 1/N .$$

Consider a sequence of alternatives $\{\delta_k\}$ which converge to the null, $\delta = 0$, and sequences of test statistics $\{Z_k\}$ and $\{Z_k^*\}$ for testing H_0 $\delta = 0$. The efficacies of Z and Z^* are given by

$$e(Z) = \frac{\frac{d}{d\delta} u(\delta)}{\sqrt{Nv^2(\delta)}} = -\bar{c}/\sigma \left[\sum_i \theta_i \phi_i \beta_i / (\theta_i + \phi_i)\right] ,$$

and

$$e(Z^*) = \frac{\frac{d}{d\delta} u^*(\delta)}{\sqrt{Nv^{*2}(\delta)}} = -\bar{c}/\sigma^* \left[\sum_i \theta_i \beta_i / 2\right] .$$

The asymptotic relative efficiency (Pitman efficiency), denoted by ARE, of matching to random sampling is therefore,

$$\text{ARE} = e(Z^*)^2/e(Z)^2 = \frac{\left(\sum_i \theta_i \beta_i/2\right)^2/\sigma^{*2}}{\left(\sum_i \theta_i \phi_i \beta_i/(\ _i + \ _i)\right)^2/\sigma^2}$$

$$= \frac{\text{"expected" M-H } X_1^2 \text{ statistic for matching}}{\text{"expected" M-H } X_1^2 \text{ statistic for random sampling}}$$

for Pitman alternatives to the null.

The regularity conditions (Gibbons, 1971) are checked below.

1. $\frac{d}{d\delta} u(\delta) = -\frac{\bar{c}}{\sigma} \sum_i \theta_i \phi_i \beta_i / (\theta_i + \phi_i)$, which is non-zero and continuous for $\delta = 0$ (recall $\bar{c} > 0$) .

$\frac{d}{d\delta} u^*(\delta) = - \frac{\bar{c}}{\sigma^*} \sum_i \theta_i \beta_i / 2$, which is also non-zero and continuous for $\delta = 0$.

2. $\lim_{N\to\infty} \frac{\frac{d}{d\delta}(u(\delta))}{\sqrt{Nv^2(\delta)}} = - \frac{\bar{c}}{\sigma} \sum_i \theta_i \phi_i \beta_i / (\theta_i + \phi_i)$, which is constant for all N .

$\lim_{N\to\infty} \frac{\frac{d}{d\delta}(u^*(\delta))}{\sqrt{Nv^{*2}(\delta)}} = - \frac{\bar{c}}{\sigma^*} \sum_i \theta_i \beta_i / 2$, constant for all N .

3. $\lim_{N\to\infty} \frac{\frac{d}{d\delta}(u(\delta)|\delta=\delta_k)}{\frac{d}{d\delta}(u(\delta)|\delta=0)} = \lim_{N\to\infty} \frac{\frac{d}{d\delta}(u^*(\delta)|\delta=\delta_k)}{\frac{d}{d\delta}(u^*(\delta)|\delta=0)} = 1$.

4. $\lim_{N\to\infty} \frac{v(\delta)|\delta=\delta_k}{v(\delta)|\delta=0} = \lim_{N\to\infty} \frac{v^*(\delta)|\delta=\delta_k}{v^*(\delta)|\delta=0} = \frac{1/N}{1/N} = 1$.

5. $\frac{Z-E(Z)}{\sigma}$ and $\frac{Z^*-E(Z^*)}{\sigma^*}$ both follow limiting standard normal distributions under the null hypothesis and Pitman alternatives.

Therefore, the regularity conditions are satisfied, and the proof is complete.

APPENDIX 3

Proposition: cORm ≤ OR, where OR = uniform, stratum-specific odds ratio, and OR > 1

cORm = crude matched odds ratio .

Proof: By the assumption of uniformity $OR = \frac{\varepsilon_i(1-\delta_i)}{\delta_i(1-\varepsilon_i)}$, ∀i. This implies that

$$\varepsilon_i = \delta_i OR/(1+\delta_i(OR-1)) ,$$

and

$$1-\varepsilon_i = (1-\delta_i)/(1+\delta_i(OR-1)) .$$

Now,

$$\begin{aligned} cORm &= \frac{\left(\sum_i \varepsilon_i v_i\right)\left(1-\sum_i \delta_i v_i\right)}{\left(\sum_i \delta_i v_i\right)\left(1-\sum_i \varepsilon_i v_i\right)} \\ &= \frac{\left[\sum_i OR\delta_i v_i/(1+\delta_i(OR-1))\right]\left(1-\sum_i \delta_i v_i\right)}{\left[\sum_i (1-\delta_i)v_i/(1+\delta_i(OR-1))\right]\left(\sum_i \delta_i v_i\right)} \\ &= OR \cdot \frac{\left(\sum_i \delta_i v_i/f_i\right)\left(1-\sum_i \delta_i v_i\right)}{\left(\sum_i (1-\delta_i)v_i/f_i\right)\left(\sum_i \delta_i v_i\right)} , \quad \text{where } f_i = 1+\delta_i(OR-1) \\ &= OR \cdot g(\underset{\sim}{v}, \underset{\sim}{\delta}, OR) . \end{aligned}$$

It is now demonstrated that $g(\cdot) \leq 1$. Assume $g(\cdot) > 1$. Then

$$\left[\sum_i \delta_i v_i/f_i\right]\left[1 - \sum_i \delta_i v_i\right] > \left(\sum_i \delta_i v_i\right)\left[\sum_i (1-\delta_i)v_i/f_i\right] ,$$

and

$$\left(\sum_i \delta_i v_i/f_i\right) > \left(\sum_i \delta_i v_i\right)\left(\sum_i v_i/f_i\right) .$$

Therefore,

$$\sum_i \frac{v_i}{f_i}\left(\delta_i - \left(\sum_j \delta_j v_j\right)\right) > 0 \ . \qquad (3.1)$$

Now

$$\begin{aligned}\delta_i - \sum_j \delta_j v_j &= \delta_i(1-v_i) - \sum_{j\neq i} \delta_i v_i \\ &= \delta_i \sum_{j\neq i} v_i - \sum_{j\neq 1} \delta_i v_i \\ &= \sum_j (\delta_i - \delta_j) v_j \ .\end{aligned}$$

And, by substituting into (3.1),

$$\sum_i \frac{v_i}{f_i}\left(\sum_j (\delta_i - \delta_j) v_j\right) > 0 \ ,$$

or

$$\sum_i \sum_j \frac{v_i v_j}{f_i} (\delta_i - \delta_j) > 0 \ . \qquad (3.2)$$

By summing over j only up to i, then (3.2) can be expressed as

$$\sum_{i>j} \sum v_i v_j \left(\frac{1}{f_i} - \frac{1}{f_j}\right)(\delta_i - \delta_j) > 0 \ . \qquad (3.3)$$

And $\frac{1}{f_i} - \frac{1}{f_j} = \frac{f_j - f_i}{f_i f_j} = \frac{(OR-1)}{f_i f_j}(\delta_j - \delta_i)$. Therefore (3.3) is written as

$$-\sum_{i>j} \sum v_i v_j \frac{(OR-1)}{f_i f_j} (\delta_i - \delta_j)^2 > 0 \ ,$$

which holds only if $OR < 1$. Since $OR > 1$ by assumption, (3.4) is false, $g(\cdot) \leq 1$, and the proof is complete. Note also that if $OR < 1$, then $cORm \geq OR$. In both cases, if the δ_i are uniform, then $cORm = OR$, and there is no confounding.