HOW CASE-MATCHING CAN REDUCE DESIGN EFFICIENCY IN RETROSPECTIVE STUDIES

BY

MYRA L. SAMUELS

TECHNICAL REPORT NO. 50
SEPTEMBER 1979

PREPARED UNDER THE AUSPICES OF
PUBLIC HEALTH SERVICE GRANT 5 RO1 GM21215-05

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
HOW CASE-MATCHING CAN REDUCE DESIGN EFFICIENCY

IN RETROSPECTIVE STUDIES

BY

MYRA L. SAMUELS

TECHNICAL REPORT NO. 50

SEPTEMBER 1979

PREPARED UNDER THE AUSPICES

OF

PUBLIC HEALTH SERVICE GRANT 5 R01 GM21215-05

DIVISION OF BIOSTATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
## Contents

1. **Introduction**  
   Page 1

2. **Framework for Comparison of Design and Analysis Strategies**  
   2.1. Population Models  
   Page 4  
   2.2. Sampling Models, Test Statistics, and Strategies  
   Page 8

3. **Model F: X Related to F but not to D**  
   3.1. An Example  
   Page 11  
   3.2. General Results for Model F  
   Page 16  
   3.3. Relation to Prospective Studies  
   Page 20  
   3.4. Practical Considerations  
   Page 23

4. **Model D: X Related to D but not to F**  
   4.1. Introductory Examples  
   Page 26  
   4.2. Results for Model D, Constant Odds Ratio Case  
   Page 29  
   4.3. The Case of Stratum-Dependent Odds Ratio  
   Page 32

5. **Computations and Proofs**  
   5.1. Basic Relations  
   Page 34  
   5.2. Proofs for Section 3.2  
   Page 41  
   5.3. Computations for Section 3.4  
   Page 42  
   5.4. Proofs for Sections 4.2 and 4.3  
   Page 43
HOW CASE-MATCHING CAN REDUCE DESIGN EFFICIENCY
IN RETROSPECTIVE STUDIES

by

Myra L. Samuels

Summary

In a retrospective, or case-control study of the relationship between a disease $D$ and a putative etiologic factor $F$, the cases and controls are often pair-matched with respect to an auxiliary variable $X$. If $X$ is not a confounding factor (i.e., if $X$ is related to either $F$ or $D$ but not to both), it is shown that such matching can be disadvantageous from an efficiency point of view. In particular, if $X$ is related to $F$ but not to $D$, the matched design is always less efficient (asymptotically, in the sense of Pitman) than the unmatched design. This finding contradicts a widely held belief based on prospective studies; the apparent paradox is resolved by noting an asymmetry in the relation between the prospective and retrospective designs. If $X$ is related to $D$ but not to $F$, matching can be advantageous in certain circumstances; however, if the incidence of $D$ is small, then matching with respect to such an $X$ cannot improve design efficiency.

Key Words and Phrases: retrospective studies, case-matching, contingency tables, confounding variables, Mantel-Haenszel tests, chi-square tests
1. Introduction

The retrospective, or case-control, study design is widely used in the investigation of the etiology of diseases such as cancer, heart disease, thromboembolic disease, etc. The essence of this design is that a number of cases (individuals who have the disease) and a number of controls (individuals free of the disease) are investigated as to their exposure to one or more putative etiologic factors. Typically, in retrospective studies the cases are relatively rare and all those available are used. On the other hand, the available pool of controls is large, so that questions are raised as to how they should be selected. A favorite device is case-matching: for each case, a control is chosen who matches the case with respect to one or more auxiliary variables such as age, sex, geographic location, etc.

The role of matching in retrospective studies has been the subject of considerable confusion in the epidemiologic and statistical literature. The claim has been made that matching is useful only with respect to confounding variables (Miettinen 1968, 1970); the counterclaim is that matching can increase the efficiency of the study (Bross 1969). Of course, the usefulness of matching as a device to increase efficiency is well-recognized among statisticians. As a consequence, statisticians addressing the subject of retrospective studies have either taken the efficiency argument for granted (Fleiss 1973, Schlesselman 1974) or have vigorously pointed out that, (in accordance with the usual statistical arguments and in contradiction to the usual epidemiologic practice) matching for efficiency should emphasize those auxiliary variables which
are strongly related to the etiologic factor rather than to the disease itself (Worcester 1964, McKinlay 1977).

It is the aim of this report to clarify the issues underlying the controversy surrounding matching, and to quantitatively characterize the advantages and disadvantages of matching in certain situations. Specifically, consider a given auxiliary variable \( X \), a given disease \( D \), and a given factor \( F \). If \( X \) is related to either \( D \) or \( F \) but not to both, then \( X \) could be ignored in the study design and analysis without affecting the validity of the conclusions. If, on the other hand, \( X \) is related both to \( D \) and to \( F \), then \( X \) is a confounding variable and the relationship between \( D \) and \( F \) can be validly studied only by taking account of \( X \) in some way.

In this report we shall investigate the effect of matching (with respect to \( X \)) on the power with which the suspected etiologic relationship between \( F \) and \( D \) can be detected, in the case where \( X \) is not a confounding variable. Issues such as the expense of matching and the imperfections of matching with respect to a continuous \( X \) will not be discussed. The focus will be on the important and generally neglected difference between the impact of matching in retrospective designs and its impact in prospective designs. We shall find that in the absence of confounding matching can reduce the asymptotic efficiency of a retrospective study design. In particular, if \( X \) is related to \( F \) but not to \( D \), the matched design is always less efficient than the unmatched design. If \( X \) is related to \( D \) but not to \( F \) matching can be advantageous in certain circumstances; however, if the incidence of \( D \) is small, then matching with respect to such an \( X \) cannot improve (and may reduce) design efficiency.
When $X$ is a confounding variable, questions arise as to whether $X$ is best controlled by matching or by post-stratification, and, in the case when the bias introduced by ignoring $X$ is conservative, whether it might be better to ignore $X$ than to match on $X$. The answers to these questions will depend upon the relationship between $X$, $F$, and $D$. The case of a confounding variable $X$ will be discussed in a future report.
2. Framework for Comparison of Design and Analysis Strategies

2.1. Population Models

In this section we describe a general framework for defining the relationship between a disease \( D \), a factor \( F \), and an auxiliary variable \( X \); within this framework the meaning of confounding can be precisely specified.

**Notation.** Consider a population of individuals at a fixed time; each individual \((i)\) either has the disease \((D)\) or is a potential control \((C)\), \((ii)\) either has been exposed to the factor \((F)\) or not \((\bar{F})\), and \((iii)\) has a value of \( X \). We assume that \( X \) is a discrete variable taking values \( x_1, \ldots, x_k \); thus \( X \) partitions the population into \( k \) homogeneous strata.

Let

\[
\begin{align*}
    p_i &= P(D|\bar{F}, X=x_i) \\
    p'_i &= P(D|F, X=x_i) \\
    q_i &= P(F|X=x_i) \\
    r_i &= P(X=x_i) ; \quad i = 1, \ldots, k
\end{align*}
\]

Thus \( \{p'_i\} \) and \( \{p_i\} \) are the stratum-specific incidence rates of \( D \) with and without exposure to \( F \), \( \{q_i\} \) are the stratum-specific exposure rates, and \( \{r_i\} \) is the distribution of the population among the strata.
Let $p$, $p'$, $q$, and $r$ represent the $(k \times 1)$ vectors of the corresponding parameters; also, let

$$p = \sum p_i r_i$$

$$p' = \sum p'_i r_i$$

(2)

$$q = \sum q_i r_i$$

be the population average incidence rates and exposure rates.

The effect of $F$ on $D$ can be expressed in terms of the incidence ratios $p'_i / p_i$ or the odds ratios

$$\theta_i = \frac{p'_i (1-p_i)}{p_i (1-p'_i)}.$$

(3)

When $D$ is rare (i.e., $p_i$ and $p'_i$ are small) the incidence ratios are nearly equal to the odds ratios.

**Definition of confounding.** In common usage, $X$ is said to be a confounding variable if ignoring $X$ "distorts" the true relationship between $D$ and $F$. Let us make this precise.

The null hypothesis that $F$ does not contribute to $D$, and corresponding one-sided alternatives, may be expressed as

$$H_0: p_i = p'_i \text{ for all } i$$

$$H_1: p_i < p'_i \text{ for all } i$$

(4)

$$H_2: p_i > p'_i \text{ for all } i$$

these may also be stated as
\( H_0: \) \( D,F \) independent, conditional on \( X \)

\( H_1: \) \( D,F \) positively dependent, conditional on \( X \) \hspace{1cm} (5)

\( H_2: \) \( D,F \) negatively dependent, conditional on \( X \)

If \( X \) is ignored, \( H_0 \) cannot be tested against \( H_1 \) or \( H_2 \); instead one must be satisfied to test the corresponding unconditional hypotheses

\[ H^*_0: \] \( D,F \) independent

\[ H^*_1: \] \( D,F \) positively dependent \hspace{1cm} (6)

\[ H^*_2: \] \( D,F \) negatively dependent.

We shall say that \( X \) is not a confounding variable if the relationship between \( X \) and \( F \) and between \( X \) and \( D \) is such that

\[ H_j = H^*_j, \hspace{0.5cm} j = 0,1,2. \] \hspace{1cm} (7)

Thus, if \( X \) is not a confounding variable, testing \( H^*_0 \) instead of \( H_0 \) cannot be qualitatively misleading: if the relationship between \( F \) and \( D \) is in the same direction in all strata, then the "apparent" relationship ignoring \( X \) will be in the same direction.

Models without confounding. There are two simple cases, here called Model \( F \) and Model \( D \), in which the auxiliary variable \( X \) is not a confounding variable:

Model \( F \). \( X \) is related to \( F \) but not to \( D \). Formally,

\[ p_i = p \hspace{1cm} \text{for all } i \] \hspace{1cm} (8)

\[ p'_i = p' \hspace{1cm} \text{for all } i \]
Model D. X is related to D but not to F. Formally,

\[ q_i = q \quad \text{for all } i \]  \hspace{1cm} (9)

In Model F the incidence rates of D depend on exposure to F but are constant over strata. For example, X might be geographic location, F might be a dietary factor, and D might be a disease (e.g., colon cancer) to which diet, but not geographic location, makes a direct etiological contribution. In Model D the incidence rates of D can have any structure, but the risk of exposure is constant over strata. For example, X might be age, D lung cancer, and F an etiological factor, (e.g., an airborne pollutant) to which old and young are exposed with equal likelihood. Of course, Models F and D are not mutually exclusive. They intersect when both (8) and (9) hold, so that X is unrelated to both F and D.

The formal definitions (8) and (9) of Models F and D are the only reasonable translations of the verbal definitions. Yet (8) and (9) are not mathematically similar: (8) states that X and D are \textit{conditionally} independent given exposure status, while (9) states that X and F are \textit{unconditionally} independent. The difference in the two translations reflects the fact that F and X are both possible \textit{causes} of D. As we shall see, this fundamental asymmetry between F and D lies at the root of much of the confusion surrounding the role of matching in retrospective studies.

In spite of the asymmetry in the definitions of Models F and D, under either model X is not a confounding variable in the sense of (7). Model D implies the relations
\[ P(D|\overline{F}) = p \]
\[ P(D|F) = p' , \]

which imply (7). Model F implies collapsibility (see Bishop, Feinberg and Holland 1975, p. 39) which guarantees that, in fact,

\[ H_j \propto H_j^* \quad j = 0,1,2 . \]

### 2.2. Sampling Models, Test Statistics, and Strategies

**Matched and unmatched sampling.** We assume that the entire population is available for sampling. In both sampling schemes, a fixed number \( n \) of cases (D) are selected at random from the diseased individuals, and an equal number of controls (C) are selected from the remainder of the population. In **unmatched sampling**, the controls are chosen at random. In **matched sampling**, as each case is chosen his stratum value \( x_i \) is noted, and a control is chosen at random from the nondiseased individuals in stratum \( x_i \). In either case exposure to \( F \) is then determined for each individual, and the data can be arranged in a series of \( k \) contingency tables. The table for the \( i \)-th stratum can be displayed as follows:

\[
\begin{array}{c|cc}
\text{Disease Status} \\
\hline
\text{F} & D & C \\
\hline
K_1 & L_1 \\
\hline
\overline{F} & N_{1i}-K_i & N_{2i}-L_i \\
\hline
N_{1i} & N_{2i} \\
\end{array}
\]

(10)
These tables can be collapsed over strata to produce the table

\[
\begin{array}{c|cc|}
\text{Disease Status} & D & C \\
\hline
F & K & L \\
\bar{F} & n-K & n-L \\
\hline
n & n
\end{array}
\]

where \( K = \Sigma K_i \), \( L = \Sigma L_i \). The marginal condition \( \Sigma N_{1i} = \Sigma N_{2i} = n \) must hold for either sampling scheme; matched sampling imposes the additional constraints \( N_{1i} = N_{2i}, i = 1, \ldots, k \).

The Mantel-Haenszel test. A popular method for combining the \( k \) contingency tables to obtain an overall test of \( H_0 \) is the Mantel-Haenszel test, based on the statistic

\[
W^2_S = \frac{(\Sigma(K_i - E_i))^2}{\Sigma V_i}
\]

where, letting \( T_i = N_{1i} + N_{2i} \),

\[
E_i = N_{1i}(K_i + L_i)/T_i
\]

and

\[
V_i = E_i N_{2i} (T_i - K_i - L_i)/T_i (T_i - 1) .
\]

Under \( H_0 \), \( W^2_S \) is asymptotically distributed as \( \chi^2_1 \). Except when the \( T_i \) are small, the Mantel-Haenszel test is essentially the same as Cochran's test (see Fleiss 1973). Birch (1964) and Radhakrishna (1965) have shown
that these tests are optimal in certain models when the $\{N_{ij}\}$ and $\{N_{2i}\}$ are regarded as fixed.

McNemar's test (see Fleiss 1973), which treats the data in its original pairs, is less powerful than the Mantel-Haenszel stratified test in the present situation (where several pairs may be expected to arise from each homogeneous stratum) and will not be further considered.

Strategies for design and analysis. Given the data, the Mantel-Haenszel test is a reasonable method of analysis (although not the only one), and we shall use it as a basis for efficiency comparisons. This leaves open the question of how the data should be generated—by matched or unmatched sampling. A further question is raised by the possibility of applying the Mantel-Haenszel test to the collapsed contingency table (11); the resulting statistic, say $W^2_C$, is essentially the same as an ordinary Pearson's chi-square statistic.

Thus, with respect to a given $X$, we shall consider four candidate strategies for the design and analysis of a retrospective study of the effect of $F$ on $D$: the data may be collected by matched or unmatched sampling, and a Mantel-Haenszel test may be performed on either the stratified or the collapsed contingency table. The relative merits of these strategies depend on the population structure with respect to $X$, $F$, and $D$. In the following sections the strategies will be examined separately for Model $F$ and for Model $D$. The basis for comparison will be the asymptotic relative efficiencies of competing valid tests of $H_0$. 
3. Model F: X related to F but not to D

It is well known that an experiment to compare two treatments can be made more efficient if experimental subjects are matched pairwise with respect to a covariate that is related to the response variable. Since the response variable in a retrospective study is exposure status, it might appear at first glance that matching should be advantageous in Model F. We shall show that in fact the reverse is true: the effect of matching in Model F is always to reduce design efficiency. Before giving general results, we illustrate this effect by an example.

3.1. An Example

Table 1 shows the population parameters (defined in Equation (1)) for an illustrative population obeying Model F, which we shall call Example 1.

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Disease Incidence</th>
<th>Exposure Probability</th>
<th>Stratum Size</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>( p_i )</td>
<td>( p'_i )</td>
<td>( q_i )</td>
<td>( r_i )</td>
</tr>
<tr>
<td>1</td>
<td>.0001</td>
<td>.0002</td>
<td>.9</td>
<td>.2</td>
</tr>
<tr>
<td>2</td>
<td>.0001</td>
<td>.0002</td>
<td>.1</td>
<td>.8</td>
</tr>
</tbody>
</table>

Collapsed: \( p_C = .0001 \) \( p'_C = .0002 \) \( \theta_C = 2.0002 \)

In this example, 20% of the population is heavily exposed (.9) to the factor F, while the remaining 80% is lightly exposed. The disease
incidence rates are stratum-independent. Consequently, the collapsed (marginal) incidence rates and odds ratios, namely

\[ p_C = P(D|\overline{F}) \]

\[ p'_C = P(D|F) \]  

(14)

and \[ \theta_C = \frac{p'_C(1-p_C)}{p_C(1-p'_C)} \]

are equal to the corresponding stratum-specific values. In other words, the population contingency table is collapsible with respect to the incidence rates, incidence ratio and odds ratio.

Table 2 shows the expected cell frequencies, stratified and collapsed, generated by unmatched sampling and by matched sampling from the population of Table 1. Each set of frequencies is based on a total of \( n=100 \) cases (D) and 100 controls (C); in the interest of clarity, the frequencies have been slightly adjusted to achieve consistent marginal totals.

Table 2. Expected Frequencies for Example 1 (n=100)

<table>
<thead>
<tr>
<th></th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>Collapsed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( D )</td>
<td>( C )</td>
<td>( D )</td>
</tr>
<tr>
<td>( F )</td>
<td>28.6</td>
<td>18.0</td>
<td>12.7</td>
</tr>
<tr>
<td>( \bar{F} )</td>
<td>1.6</td>
<td>2.0</td>
<td>57.1</td>
</tr>
<tr>
<td>( F )</td>
<td>30.2</td>
<td>20.0</td>
<td>69.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100.0</td>
</tr>
</tbody>
</table>
Table 2. Expected Frequencies for Example 1 (n=100) (continued)

**Matched Design**

<table>
<thead>
<tr>
<th></th>
<th>$x_1$</th>
<th></th>
<th>$x_2$</th>
<th></th>
<th>Collapsed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D</td>
<td>C</td>
<td>D</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>28.6</td>
<td>27.2</td>
<td>F</td>
<td>12.7</td>
<td>6.9</td>
</tr>
<tr>
<td>F</td>
<td>1.6</td>
<td>3.0</td>
<td>F</td>
<td>57.1</td>
<td>62.9</td>
</tr>
<tr>
<td></td>
<td>30.2</td>
<td>30.2</td>
<td></td>
<td>69.8</td>
<td>69.8</td>
</tr>
</tbody>
</table>

The most obvious impact of matching is on the expected frequencies among the controls. The column total 20.0 has its natural value $20.0 = nP(X=x_1|C)$ in the unmatched design, while in the matched design the corresponding total is forced to agree with $30.2 = nP(X=x_1|D)$. These expected totals would be equal only if $F$ had no effect on $D$. Within each stratum the exposure distribution among the controls is not distorted by matching, so that

$$P(F|C, X=x_1) = 18.0/20.0 = 27.2/30.2$$

and

$$P(F|C, X=x_2) = 8.0/80.0 = 6.9/69.8$$

(The numerical equalities in (15) are not exact due to round-off error.)

The collapsed table is, however, distorted by matching: in the unmatched collapsed table we have $P(F|C) = 26.0/100.0$, while the corresponding ratio $34.1/100.0$ in the matched table is not the same, and in fact cannot be interpreted as an exposure probability at all.
This distortion of the matched collapsed table is reflected in the odds ratios. It is well known (see, for example, Fleiss 1973) that if the entries in a contingency table correspond to binomial expected values, then the cross-product ratio is equal to the odds ratio, regardless of the sample sizes. For example, the cross product ratios \((28.6)(2.0)/(1.6)(18.0)\) and \((28.6)(3.0)/(1.6)(27.2)\) are both (except for round-off error) equal to \(\theta_1 = 2.0002\). These ratios are stratum-independent in either design. Further, in the unmatched design they are collapsible: the cross-product ratio in the unmatched collapsed table is also equal to \(2.0002\). The distortion induced by matching destroys this collapsibility: the cross-product ratio in the matched collapsed table is \(1.36!\)

As an indicator of the relative efficiencies of design/analysis strategies, we can compare the values of Mantel-Haenszel test statistics computed as if the expected frequencies in Table 2 were observed frequencies. The statistics can be computed from the stratified tables \((W_S^2)\) or the collapsed tables \((W_C^2)\). The results are

Unmatched Design: \(W_S^2 = 2.50\quad W_C^2 = 5.22\) \hspace{1cm} (16)

Matched Design: \(W_S^2 = 2.41\quad W_C^2 = 1.09\)

It is not surprising that the distortion of the collapsed odds ratio induced by matching shrinks the value of \(W_C^2\) for the matched design. This shrinkage has been interpreted by several authors as demonstrating the advisability of matching. In fact, it indicates only that, if matching is used in the design but not retained in the analysis, then the significance test is not valid.
The other three $W^2$ values in (16) represent valid significance tests, but are unequal, although computed from tables with the same cross-product ratio. We shall see that these three $W^2$ values can be interpreted as measures of relative design efficiency. Note that the largest $W^2$ value (5.22) results from ignoring $X$; i.e., neither matching on $X$ nor stratifying on $X$. In this example, then, matching on $X$ reduces design efficiency by about 50%.
3.2. General results for Model F

In this section we describe and interpret the effect of matching in Model F. Proofs are given in Section 5.

Effect of matching on the odds ratio. Under Model F, the odds ratios are always stratum-independent and collapsible; i.e., $\theta_i = \theta_c = \theta$ for all $i$. The observed stratum-specific cross-product ratios

$$K_i(N_{2i}-L_i)/L_i(N_{1i}-K_i)$$

are consistent (as $n \to \infty$) estimators of $\theta$ under either matched or unmatched sampling. The observed collapsed cross-product ratio

$$K(n-L)/L(n-K)$$

is a consistent estimator of $\theta$ under unmatched sampling only. Under matched sampling, (18) is a consistent estimator of a parameter, say $\theta'$, which is always shifted toward the null hypothesis: i.e., $(\theta'-1)$ has the same sign as $(\theta-1)$ but smaller absolute value (strictly smaller unless $\theta=1$ or unless $q_1=q$ for all $i$).

Appropriate criteria for efficiency comparisons. If only the analysis, and not the design, of the retrospective study is under consideration, it is reasonable to study the distributions of competing test statistics conditional on the column marginals $\{N_{1i}, N_{2i}\}$ of the tables (as have Birch 1964, Radhakrishna 1965 and Gart 1971). In the present case, we wish to compare two designs—matched and unmatched—which differ in their effect on these marginals. Consequently, we shall consider the unconditional distributions of the $W^2$ statistics, with only $\Sigma N_{1i} = \Sigma N_{2i} = n$ taken as fixed.
Asymptotic distributions of $W^2$ statistics under $H_0$. Under unmatched sampling, both $W_S^2$ and $W_C^2$ are asymptotically distributed as $\chi_1^2$ under $H_0$. Under matched sampling, $W_S^2$ is also asymptotically $\chi_1^2$ under $H_0$, but $W_C^2$ is not. In fact, under matching $W_C^2$ is strictly less than $W_S^2$ with very high probability, whether $H_0$ is true or false. Thus the collapsed Mantel-Haenzel test in the matched design is not a valid test -- it rejects $H_0$ too infrequently even if $H_0$ is true.

Asymptotic relative efficiencies of valid $W^2$ tests. The three valid $W^2$ statistics--$W_S^2$ and $W_C^2$ under unmatched sampling and $W_S^2$ under matched sampling--have different asymptotic distributions when $H_0$ is false. These distributions can be compared using asymptotic relative efficiencies (ARE's) defined as follows: Let $\omega_S^2$ and $\omega_C^2$ denote quantities calculated analogously to $W_S^2$ and $W_C^2$, but with the random variables $\{K_1, L_1, N_{11}, N_{21}\}$ replaced by their expected values under a given sampling scheme (thus the $W^2$ values given in (16) for Example 1 are actually $\omega^2$ values). In order to compare any two tests, say I and II, we can consider the ratio $\omega_I^2/\omega_{II}^2$, which depends on the population parameters $\pi, \pi', \pi$, and $\rho$ but not on $n$. The ARE is then

$$\text{ARE}[I/II] = \lim_{n \to \infty} \frac{\omega_I^2}{\omega_{II}^2}$$

(19)

where the limit is taken as $\pi' + \pi$ with $\pi$ and $\rho$ fixed. The ARE (19) is the Pitman efficiency of test I with respect to test II, and can be interpreted as the limiting value.
\[
\lim \frac{n_{II}}{n_I},
\]

(20)

where \( n_I \) and \( n_{II} \) are the respective sample sizes required for tests I and II respectively to achieve a given asymptotic power against a sequence of alternatives converging to \( H_0 \) (see Noether 1955).

Under Model F, the ARE's are

\[
\text{ARE}[w_S^2(\text{matched})/w_S^2(\text{unmatched})] = 1
\]

(21)

and

\[
\text{ARE}[w_S^2(\text{matched})/w_C^2(\text{unmatched})] = 1 - v
\]

(22)

where

\[
v = \frac{\sum r_i(q_i - q)^2}{q(1-q)}
\]

(23)

is a measure of the inter-stratum variability of exposure. The parameter \( v \) varies between 0 and 1 and can be interpreted as the correlation between the responses (\( F \) or \( \bar{F} \)) of two individuals matched by stratum.

The relations (21) and (22) indicate that (i) the stratified statistics \( w_S^2 \) are equally efficient in the matched and unmatched designs, but (ii) they are both inferior to the collapsed statistic \( w_C^2 \) in the unmatched and (iii) the greater the inter-stratum variability of exposure, the more inferior is the matched design.

The results (16) for Example 1 can serve as a numerical example of the ARE's: \( 2.41/2.50 \approx 1 \), as predicted by (21), and \( 2.41/5.22 = 0.46 \),
while (22) predicts 0.47 for this value. (The discrepancies from (21) and (22) would be even smaller if \( \theta \) were smaller). Thus, with the interpretation (20), for the population of Example 1 the matched design requires more than twice the sample size as the unmatched design in order to achieve the same asymptotic power.
3.3. Relation to Prospective Studies

The conclusions of Section 3.2 appear to contradict many of the discussions of matching and design efficiency which have appeared in the statistical literature. Such discussions invariably make the assumption (usually implicit) that the matching factor $X$ is not a confounding factor, so that either the matched or unmatched design would be valid. The only design/analysis strategies considered, then, are (i) match on $X$ and analyze as matched; and (ii) ignore $X$. The usual conclusion is that if $X$ is related to the response variable, then strategy (i) is more efficient in large samples. (Loss of efficiency due to loss of d.f. in small samples is irrelevant to the issue here considered.)

Now in a prospective study design the groups of individuals to be compared are defined by their exposure status, while the response variable is their disease status. This interchange of the roles of $F$ and $D$ appears to be the only formal difference between the prospective and retrospective designs. Since the hypotheses $H_j$ and $H_j^*$ are symmetric in $F$ and $D$, it would appear that the conclusions concerning the superior efficiency of strategy (i) should apply equally to retrospective and prospective designs. Indeed, Worcester (1964) gives quantitative descriptions of the increase in efficiency due to matching when the response variable is dichotomous; no distinction between prospective and retrospective studies is made in the calculations. Similarly, McKinlay (1977) describes both types of studies in the same notation and emphasizes that, in order for matching to be effective in increasing efficiency, the matching covariate $X$ should be related to the response variable, which in retrospective studies is exposure status.

20
The resolution of this apparent contradiction lies in the following deceptively simple fact which apparently has escaped the notice of most statisticians: the interchange of the roles of $F$ and $D$ is not the only formal difference between the prospective and retrospective designs. A second difference is to be found in the formal statement of the assumption that $X$ is not a confounding factor. Let us illustrate this with an example.

Consider a prospective study to compare colon cancer ($D$) incidence in two groups of subjects: Group $F$ consumes a low-fiber diet and Group $\bar{F}$ consumes a high-fiber diet. Consider the possibility of matching subjects by age ($X$), which is strongly related to $D$. The assumption that age is not a confounding factor is the assumption that the age distributions in Groups $F$ and $\bar{F}$ are the same. Formally, this assumption states that $X$ and $F$ are independent. Such an assumption is in fact implicitly used in calculations such as those of Worcester (1964). Under this assumption, it does indeed follow that matching increases large-sample design efficiency.

By contrast, in designing a retrospective study of dietary habits and colon cancer, we might consider a matching variable $X$ such as ethnic group, which is strongly related to dietary habits. The assumption that ethnic group is not a confounding factor, i.e., that it is not also related to colon cancer, is now not equivalent to the assumption that $X$ and $D$ are independent; in fact, if $X$ influences $F$, and $F$ influences $D$, it is quite impossible that $X$ and $D$ should be independent. The appropriate formal assumption is, rather, that $X$ and $D$ are conditionally independent given $F$ or $\bar{F}$. Under this assumption, as we have seen, matching can only reduce design efficiency.
There is, then, a fundamental lack of symmetry in the relation between the prospective and retrospective designs. Calculations intended to increase our understanding of the role of matching, either in the absence of confounding or in its presence, are only relevant insofar as this asymmetry is taken into account.
3.4. Practical Considerations

The conventional wisdom is that the effects of matching tend to be relatively small when the response variable is dichotomous. In this section we present quantitative examples of efficiency loss and distortion due to matching in Model F. These results bear out the expectation that the exposure rates must vary considerably among strata in order for matching to have much (negative) impact.

The most extreme inter-stratum exposure variability occurs when all the \( \{q_i\} \) are either 0 or 1. The parameter \( v \) defined by (23) is then equal to 1, so that the ARE of the matched vs. the unmatched design given by (22) is zero. In order to examine less extreme cases, we will consider classes of exposure distributions in which the \( \{q_i\} \) are bounded away from 0 and 1. For any values \( q^{(1)} \) and \( q^{(2)} \), let \( v^* \) be the maximal value of \( v \) among all exposure distributions \( (q, r) \) for which the exposure probabilities \( \{q_i\} \) all lie between \( q^{(1)} \) and \( q^{(2)} \). We define the extremal distribution associated with \( q^{(1)} \) and \( q^{(2)} \) to be the distribution \( (q^*, r^*) \) which achieves \( v^* \); without loss of generality the extremal distribution can be defined on only two strata.

The asymptotic relative efficiency of matched with respect to unmatched sampling given by (22) depends strongly on the extremality of the exposure distribution. Table 3 shows values of the ARE for selected extremal distributions.
Table 3. ARE of Matched vs. Unmatched Design for Selected Extremal Distributions

<table>
<thead>
<tr>
<th>q(1)</th>
<th>q(2)</th>
<th>ARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>.05</td>
<td>.95</td>
<td>.19</td>
</tr>
<tr>
<td>.1</td>
<td>.9</td>
<td>.36</td>
</tr>
<tr>
<td>.2</td>
<td>.8</td>
<td>.64</td>
</tr>
<tr>
<td>.4</td>
<td>.6</td>
<td>.96</td>
</tr>
<tr>
<td>.1</td>
<td>.5</td>
<td>.80</td>
</tr>
<tr>
<td>.1</td>
<td>.2</td>
<td>.98</td>
</tr>
<tr>
<td>.5</td>
<td>.9</td>
<td>.80</td>
</tr>
</tbody>
</table>

Table 4 shows some other quantities of interest for selected extremal distributions and selected values of the odds ratio \( \theta \). All the calculations represented in Table 4 are approximations based on the assumption that the disease \( D \) is rare (i.e., \( p \) and \( p' \) have been passed to zero). The fourth column of Table 4 shows the relative incidence

\[
\frac{P(D|X = x^{(2)})}{P(D|X = x^{(1)})}
\]

of \( D \) in the two strata; this quantity indicates how strongly the extremality of the exposure distribution is reflected in the stratum-specific incidence rates. The fifth and sixth columns describe the distortions caused by collapsing over strata in the matched design. The fifth column shows the distorted odds ratio \( \theta' \) (the stochastic limit of (18)) as a proportion of the true odds ratio \( \theta \). The sixth column shows the stochastic limit of \( W_c^2/W_s^2 \) as \( n \to \infty \) which indicates the shrinkage of the \( W^2 \) statistic induced by collapsing (even under \( H_0 \)).
Table 4. Various Properties of Selected Extremal Distributions Calculated assuming $D$ is rare

<table>
<thead>
<tr>
<th>$q(1)$</th>
<th>$q(2)$</th>
<th>$\theta$</th>
<th>Relative Incidence ($\theta^{(24)}$)</th>
<th>Matched Design $\theta'/\theta$</th>
<th>$\lim \frac{W_C^2}{W_S^2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.9</td>
<td>2</td>
<td>1.7</td>
<td>.65</td>
<td>.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>2.8</td>
<td>.45</td>
<td>.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\infty$</td>
<td>9</td>
<td>.22</td>
<td>.82</td>
</tr>
<tr>
<td>.2</td>
<td>.8</td>
<td>2</td>
<td>1.5</td>
<td>.79</td>
<td>.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>2.1</td>
<td>.64</td>
<td>.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\infty$</td>
<td>4</td>
<td>.47</td>
<td>.89</td>
</tr>
<tr>
<td>.4</td>
<td>.6</td>
<td>2</td>
<td>1.1</td>
<td>.97</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>1.3</td>
<td>.95</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\infty$</td>
<td>1.5</td>
<td>.92</td>
<td>.99</td>
</tr>
<tr>
<td>.1</td>
<td>.5</td>
<td>2</td>
<td>1.4</td>
<td>.86</td>
<td>.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>1.9</td>
<td>.73</td>
<td>.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\infty$</td>
<td>5</td>
<td>.50</td>
<td>.96</td>
</tr>
<tr>
<td>.1</td>
<td>.2</td>
<td>2</td>
<td>1.1</td>
<td>.98</td>
<td>.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>1.2</td>
<td>.96</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\infty$</td>
<td>2</td>
<td>.88</td>
<td>1.00</td>
</tr>
<tr>
<td>.5</td>
<td>.9</td>
<td>2</td>
<td>1.3</td>
<td>.89</td>
<td>.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>1.5</td>
<td>.82</td>
<td>.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\infty$</td>
<td>1.8</td>
<td>.75</td>
<td>.92</td>
</tr>
</tbody>
</table>
4. Model D: X related to D but not to F

Model D has no analog in randomized experimentation; consequently, there is no conventional wisdom as to whether matching should increase or decrease design efficiency in this model. Bross (1969) has claimed that "strong etiologic factors often mask the effects of weaker factors," and that "matching out" a strong factor can "improve the chances of detecting a real—but relatively weak—relationship in a secondary factor". On the other hand, Miettinen (1970) asserts that "When the matching factor is unrelated to exposure, no gain or loss of...validity or efficiency is possible through matching; matching is futile". We shall see that each of these contradictory assertions has some merit; however, if the disease under study is rare, then matching is not only futile but can actually reduce efficiency. We begin with two examples and then give general results. Proofs are given in Section 5.

4.1. Introductory Examples

Recall that Model D is defined by the requirement expressed in (9) that the exposure probability $q$ be stratum-independent. The structure of the relationship between the disease probability vectors $p$ and $p'$ is not specified by the model. We shall consider two examples in which $p$ and $p'$ are constrained by the requirement that the odds ratio (3) be stratum-independent. Tables 5 and 6 show the population parameters for these two examples. Examples 2 and 3 both involve a study factor $F$ whose effect on the disease $D$ is relatively weak compared to that of the stratum variable $X$. In Example 2, $D$ has high incidence, while in Example 3 $D$ is rare. Note that the odds ratios, while stratum-independent, are not collapsible in Example 2. This shrinking of the odds ratio from $\theta = 2.25$ to
$\theta_c = 1.51$, which would be reflected in the unmatched samples, may represent the "masking" effect mentioned by Bross (1969). This effect is negligible in Example 3.

Table 5. Population Parameters for Example 2

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Disease Incidence</th>
<th>Exposure Probability</th>
<th>Stratum Size</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>$p_i$</td>
<td>$p'_i$</td>
<td>$q_i$</td>
<td>$r_i$</td>
</tr>
<tr>
<td>1</td>
<td>.80</td>
<td>.90</td>
<td>.2</td>
<td>.4</td>
</tr>
<tr>
<td>2</td>
<td>.10</td>
<td>.20</td>
<td>.2</td>
<td>.6</td>
</tr>
<tr>
<td>Collapsed: $p_c = .38$</td>
<td>$p'_c = .48$</td>
<td></td>
<td>$\theta_c = 1.51$</td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Population Parameters for Example 3

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Disease Incidence</th>
<th>Exposure Probability</th>
<th>Stratum Size</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>$p_i$</td>
<td>$p'_i$</td>
<td>$q_i$</td>
<td>$r_i$</td>
</tr>
<tr>
<td>1</td>
<td>.0100</td>
<td>.0198</td>
<td>.2</td>
<td>.4</td>
</tr>
<tr>
<td>2</td>
<td>.0002</td>
<td>.0004</td>
<td>.2</td>
<td>.6</td>
</tr>
<tr>
<td>Collapsed: $p_c = .0041$</td>
<td>$p'_c = .0082$</td>
<td></td>
<td>$\theta_c = 1.989$</td>
<td></td>
</tr>
</tbody>
</table>

For each of these examples, $\omega^2$ values ($\hat{\omega}^2$ values based on expected frequencies, as defined in Sec. 3.2) have been computed, based on $n=100$. 

27
The results for Example 2 are

Unmatched Design: \[ \omega_S^2 = 2.77 \quad \omega_C^2 = 1.36 \]  
Matched Design: \[ \omega_S^2 = 4.57 \quad \omega_C^2 = 4.53 \]  

(25)

These \( \omega^2 \) values show that, for Example 2, (i) collapsing in the matched design has virtually no effect on the \( \omega^2 \) value, and (ii) the matched design is substantially superior to the unmatched design. The \( \omega^2 \) results for Example 3 are

Unmatched Design: \[ \omega_S^2 = 2.69 \quad \omega_C^2 = 4.46 \]  
Matched Design: \[ \omega_S^2 = 4.52 \quad \omega_C^2 = 4.52 \]  

(26)

Thus for Example 3 the conclusions (i) and (ii) above hold, but the superiority of the matched design is negligibly small.
4.2. Results for Model D, Constant Odds Ratio Case

In this section we give general results for Model D under the assumption that the odds ratio \( \theta \) is stratum-independent. It will be seen that the matched design is superior, but that this superiority is negligible if the disease under study is rare.

**Effect of matching on the odds ratio.** The observed stratum-specific odds ratios (17) are consistent estimators of \( \theta \) under either matched or unmatched sampling. The observed collapsed odds ratio (18) is a consistent estimator of \( \theta_c \) (defined in (14)) in the unmatched design; the stochastic limit \( \theta' \) of (18) in the matched design is in general different from \( \theta \) and \( \theta_c \). The shrinkage of \( \theta_c \) toward 1 which was observed in Example 2 represents a general phenomenon; in fact \( \theta_c - 1 \) always has the same sign as \( \theta - 1 \) but smaller absolute value. Near \( H_0 \) we have

\[
\theta_c = \theta - (\theta-1)v_p + o(\theta-1)
\]

where

\[
v_p = \frac{\sum_i (p_i - p)^2}{p(1-p)}
\]

is a measure of inter-stratum variability of risk analogous to the measure \( v \) (23) for exposure. The matched design tends to counteract the shrinkage of \( \theta_c \) toward 1. In particular,

\[
\theta' = \theta + o(\theta-1)
\]

so that, at least near \( H_0 \), collapsing over strata "masks" the effect of F less in the matched than in the unmatched design.
Asymptotic relative efficiencies of $W^2$ tests. In Model D both $W_S^2$ and $W_C^2$ are asymptotically distributed as $X_1^2$ under $H_0$, under either matched or unmatched sampling. The power of the four valid $X^2$ tests can be compared using the asymptotic relative efficiencies introduced in Section 3.2. Let

$$\pi_i = \frac{p_i}{P}$$  \hspace{1cm} (30)

and define the function $f$ by

$$f(x) = \frac{2x(1-px)}{(1+x-2px)}.$$  \hspace{1cm} (31)

Then the ARE of the two stratified tests is

$$\text{ARE}[W_S^2(\text{matched})/W_S^2(\text{unmatched})] = \left[\sum \pi_i f(\pi_i)\right]^{-1};$$  \hspace{1cm} (32)

it follows that

$$\text{ARE}[W_S^2(\text{matched})/W_S^2(\text{unmatched})] \geq 1.$$  \hspace{1cm} (33)

The ARE of the two collapsed tests is

$$\text{ARE}[W_C^2(\text{matched})/W_C^2(\text{unmatched})] = (1-v_p)^{-2},$$  \hspace{1cm} (34)

where $v_p$ is defined in (28); it follows that

$$\text{ARE}[W_C^2(\text{matched})/W_C^2(\text{unmatched})] \geq 1.$$  \hspace{1cm} (35)

Within the matched design,

$$\text{ARE}[W_S^2(\text{matched})/W_C^2(\text{matched})] = 1;$$  \hspace{1cm} (36)
the corresponding ARE within the unmatched design may be greater or less than one, depending on the other parameters.

Thus the matched design has twin advantages in this model: (i) the tables can be collapsed over strata with minimal effect on the odds ratio and no effect on the asymptotic power of the hypothesis test, and (ii) the hypothesis tests in the matched design are at least as efficient than the corresponding tests in the unmatched design.

The case of a rare disease. Most diseases studied by the retrospective method are fairly rare. This can be expressed in the model by letting \( p \to 0 \) with \( \gamma = \frac{p}{\hat{p}} \) held fixed. In this limiting situation, the response rates \( P(F|D, X=x_i) \) and \( P(F|C, X=x_i) \) are stratum-independent, and \( \theta_c = \theta \). Further \( \nu_p = 0 \), so that (35) becomes an equality; in fact, in the limiting model \( \omega^2_{S}(\text{matched}) = \omega^2_{C}(\text{matched}) = \omega^2_{C}(\text{unmatched}) \). (37)

The inequality (33) remains strict even when \( p \to 0 \) (unless \( p_i/p = 1 \) for all \( i \)).

Thus the advantages of matching in Model D vanish in the case of a rare disease: in either the matched or the unmatched design the odds ratio is collapsible; the collapsed tests based on \( W^2_{C} \) are equally efficient in the two designs and each is at least as efficient as the stratified test based on \( W^2_{S} \).
4.3. The Case of Stratum-Dependent Odds Ratio

The one-sided alternative hypotheses \( H_1 \) and \( H_2 \) specify that the sign of \( (\theta_i - 1) \), but not its value, be stratum-independent. The description given in Section 4.2 of the effect of matching in Model D is not entirely accurate under this weaker assumption.

The relation (36) holds regardless of variation in the odds ratio. However, each of the inequalities (33) and (35) can reverse if the odds ratio is appropriately stratum-dependent. This effect is illustrated by the population of Example 4, described in Table 7. The \( \omega^2 \) values for

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Disease Incidence</th>
<th>Exposure Probability</th>
<th>Stratum Size</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( p_i )</td>
<td>( p_i' )</td>
<td>( q_i )</td>
<td>( r_i )</td>
</tr>
<tr>
<td>1</td>
<td>.05</td>
<td>.20</td>
<td>.2</td>
<td>.5</td>
</tr>
<tr>
<td>2</td>
<td>.50</td>
<td>.54</td>
<td>.2</td>
<td>.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collapsed: ( p_c = .275 ) ( p_c' = .37 ) ( \theta_c = 1.55 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Example 4, based on \( n=100 \), are

Unmatched Design: \[ \omega_S^2 = 3.07 \quad \omega_C^2 = 1.59 \]

Matched Design: \[ \omega_S^2 = 1.31 \quad \omega_C^2 = 1.29 \]

It is seen that for this population the unmatched design is strictly superior.
The case of a rare disease. In the limiting model with $p \to 0$, (37) holds regardless of variation in the odds ratio, although the inequality (33) can reverse. Thus, if the disease under study is rare, the matched design is never advantageous in Model D; if the odds ratio is suitably stratum-dependent, it can be strictly disadvantageous.
5. Computations and Proofs

5.1. Basic Relations

Definition of contingency table parameters. The following parameters will be useful for expressing the expected frequencies in the contingency tables. Let

\[ d_i = P(F|D, X=x_i) \]
\[ c_i = P(F|C, X=x_i) \]
\[ \hat{d}_i = P(X=x_i|D) \]
\[ \hat{c}_i = P(X=x_i|C) \] \hspace{1cm} (A1)

and

\[ \hat{d} = \sum_i \hat{d}_i \]
\[ \hat{c} = \sum_i \hat{c}_i \]
\[ \hat{c} = \sum_i \hat{c}_i \] \hspace{1cm} (A2)

also, let

\[ \hat{d} = \sum_i d_i \hat{d}_i = P(F|D) \]
\[ \hat{c} = \sum_i c_i \hat{c}_i = P(F|C) \]

Relation of contingency table parameters to population parameters. Let

\[ \Delta_i = (p_i^l - p_i)/p_i \] \hspace{1cm} (A3)

Straightforward computation yields the following relations between the parameters defined in (A1) and the underlying population parameters (1):

\[ d_i = (1 + \Delta_i)q_i/(1 + \Delta_i q_i) \]
\[ c_i = q_i[1 - p_i(1 + \Delta_i)]/[1 - p_i(1 + \Delta_i q_i)] \]
\[ \hat{d}_i = r_i p_i(1 + \Delta_i q_i)/[\sum r_j p_j(1 + \Delta_j q_j)] \] \hspace{1cm} (A4)
\[ \hat{c}_i = r_i[1 - p_i(1 + \Delta_i q_i)]/[\sum r_j[1 - p_j(1 + \Delta_j q_j)]] \] .
The stratum-wise null hypothesis \( H_0 \) of (5) can be restated as

\[
H_0: \Delta_i = 0 \text{ for all } i \quad (A5)
\]

or as

\[
H_0: d_i = c_i \text{ for all } i. \quad (A6)
\]

The collapsed null hypothesis \( H_0^* \) of (6) can be restated as

\[
H_0^*: \hat{d} = \bar{c}. \quad (A7)
\]

**Expected contingency table frequencies.** In the unmatched design, the expected values of (10) and (11) are

**UNMATCHED DESIGN**

\[
\begin{array}{ccc}
& D & C \\
F & \frac{nd_i f_i}{n f_i} & \frac{nc_i \tilde{f}_i}{n \tilde{f}_i} \\
\bar{F} & \frac{n(1-d_i)\tilde{f}_i}{n \tilde{f}_i} & \frac{n(1-c_i)\tilde{f}_i}{n \tilde{f}_i}
\end{array}
\]

\[
\begin{array}{ccc}
& D & C \\
F & n\hat{d} & n\bar{c} \\
\bar{F} & n(1-\hat{d}) & n(1-\bar{c})
\end{array}
\]  

(A8)

The expected contingency table frequencies in the matched design are

**MATCHED DESIGN**

\[
\begin{array}{ccc}
& D & C \\
F & \frac{nd_i f_i}{nf_i} & \frac{nc_i \tilde{f}_i}{nf_i} \\
\bar{F} & \frac{n(1-d_i)\tilde{f}_i}{nf_i} & \frac{n(1-c_i)\tilde{f}_i}{nf_i}
\end{array}
\]

\[
\begin{array}{ccc}
& D & C \\
F & n\hat{d} & n\bar{c} \\
\bar{F} & n(1-\hat{d}) & n(1-\bar{c})
\end{array}
\]  

(A9)
Distributions of contingency table entries. Let \( K = (k_1, \ldots, k_k)' \), \( L = (l_1, \ldots, l_k)' \), \( \mathbf{n}_1 = (n_{11}, \ldots, n_{1k})' \), and \( \mathbf{n}_2 = (n_{21}, \ldots, n_{2k})' \) be the vectors of the contingency table frequencies (10). The following lemmas are easily established.

**Lemma 1.**

(a) Under either matched or unmatched sampling,

(i) the \( (k \times 4) \) data matrix \( (K, L, n_1, n_2) \) is the sum of \( n \) independent and identically distributed matrices; and

(ii) conditional on \( n_1 \) and \( n_2 \), \( K_1 \) and \( L_1 \) are independent binomial random variables with respective parameters \( (n_{11}, d_1) \) and \( (n_{21}, c_1) \).

(b) Under unmatched sampling, \( n_1 \) and \( n_2 \) are independent multinomial vectors with respective parameters \( (n, \tilde{p}_1, \ldots, \tilde{p}_k) \) and \( (n, \tilde{q}_1, \ldots, \tilde{q}_k) \).

(c) Under matched sampling, \( n_1 \) is a multinomial vector with parameter \( (n, \tilde{p}_1, \ldots, \tilde{p}_k) \), and \( n_1 = n_2 \).

**Lemma 2.**

(a) Under unmatched sampling, \( K \) and \( L \) are independent binomial random variables with respective parameters \( (n, \hat{d}) \) and \( (n, \hat{c}) \).

(b) Under matched sampling, \( K \) and \( L \) are binomial random variables with respective parameters \( (n, \hat{d}) \) and \( (n, \hat{c}) \); further, \( K \) and \( L \) are independent if

\[
d_i = d \text{ for all } i \tag{A10}
\]

and

\[
c_i = c \text{ for all } i .
\]
Asymptotic null distribution of $W^2$ statistics. The null distribution of Mantel-Haenszel statistics is well known, but will be derived here for clarity of exposition.

Theorem 1. As $n \to \infty$, the following distributions are asymptotically $\chi^2_1$ under $H_0$:

(a) the distribution of $W_S^2$ under either matched or unmatched sampling; and

(b) the distribution of $W_C^2$

(i) under unmatched sampling in either Model F or Model D, and

(ii) under matched sampling in Model D.

Proof. Let

$$M_i = (N_{1i}, N_{2i}, K_i + L_i, T_i - K_i - L_i).$$  \hspace{1cm} (A11)

Then, conditionally on $(M_1, ..., M_k)$, it follows from (A6) and Lemma 1(a)(ii) that under $H_0$ the $\{K_i\}$ are (under either sampling scheme) independent hypergeometric random variables with means and variances $\{E_i\}$ and $\{V_i\}$ given by (13). It follows that $W_S^2$ is (conditional on $(M_1, ..., M_k)$ and therefore also unconditionally) asymptotically $\chi^2_1$ under $H_0$, and (a) is established. Lemma 2(a), the relation (7), and (A7) imply that the null distribution of $K$ conditional on $(M_1, ..., M_k)$ is hypergeometric, and (b)(i) is established by an argument similar to the above. Another similar argument establishes (b)(ii) after noting from (A4) and (A5) that, under Model D, $H_0$ implies (A10).

Asymptotic distribution of $W^2$ under alternatives. The notion of asymptotic relative efficiency involves a sequence of alternatives converging to $H_0$: In order to apply this idea to the vector null hypothesis in (A5), let
$\Lambda_1 = \Lambda \delta_1$ \hspace{1cm} (A12)

where $\sum r_i \delta_i = 1$; the null hypothesis can be taken to be

$H_0: \Lambda = 0$ \hspace{1cm} (A13)

with $p, q, r$, and $\delta = (\delta_1, \ldots, \delta_k)'$ regarded as nuisance parameters.

For each design and analysis scheme under consideration, the $W^2$ computed from expected values may be written as

$$W^2 = w^2(n, \Lambda; \delta, p, q, r)$$

to show explicitly its dependence on the parameters and sample size. The interpretation (20) for (19) follows from the following theorem (see Noether 1955).

Theorem 2. As $n \to \infty$ with $\delta, p, q, r$ fixed, suppose that $\Lambda^{(n)} \to 0$ in such a way that

$$w(n, \Lambda^{(n)}; \delta, p, q, r) \to L < \infty;$$

then the distributions listed in Theorem 1 are asymptotically noncentral $\chi^2$ with one d.f. and noncentrality parameter $L$.

Proof. We consider the stratified analysis first. Equations (13) show $E_i$ and $V_i$ as functions of the margins $\sum_i$; let us explicitly denote these functions by $e(\cdot)$, and $v(\cdot)$, respectively, so that

$$W^2_S = \frac{\sum (K_i - e(M))}{v(M_i)}.$$ \hspace{1cm} (A14)

Let $\kappa_i$ and $\mu_i$ denote the expected values of $K_i$ and $\sum_i$, computed at an arbitrary parameter point $(\Lambda, \delta, p, q, r)$; the value of $\mu_i$
depends upon whether the sampling design is matched or unmatched. Then \( \omega^2 \) can be written as

\[
\omega^2 = \frac{\sum \left( \kappa_i - e(\mu_i) \right)^2}{\Sigma \nu(\mu_i)}.
\] (A15)

Because of Lemma 1 and the central limit theorem, the quantity

\[ n^{-1/2} \sum \left( [K_i - e(M_i)] - [\kappa_i - e(\mu_i)] \right) \] is asymptotically normal; let \( \sigma^2 \)
de note its asymptotic variance. Then \( W^2 \) can be written as

\[
W^2 = B_n [A Z_n + \omega_n]^2 \] (A16)

where

\[
A^2 = n \sigma^2 / \Sigma \nu(\mu_i)
\]

\[
B_n = \Sigma \nu(\mu_i) / \Sigma \nu(M_i)
\]

and

\[
Z_n = \sum \left( [K_i - e(M_i)] - [\kappa_i - e(\mu_i)] \right) / \sigma \sqrt{n}.
\]

As \( n \to \infty \) with all parameters fixed, \( Z_n \sim h(0, 1) \); also, it is clear from
the form of the function \( \nu(\cdot) \) and from Lemma 1 that \( A \) does not depend
upon \( n \) and that \( B^n \sim P \). Letting \( n \Sigma \)
denote the covariance matrix of the vector of observed frequencies in the entire \( 2 \times 2 \times k \) contingency table,
we see from Lemma 1 and (A4) that \( \Sigma \) is a bounded and continuously differentiable function of \( \Delta \). Consider now a sequence \( \Delta^{(n)} \to 0 \) as specified
in the theorem. Because \( \sigma \) is a continuous function of \( \Delta \), it follows
from Theorem 1 that \( A \to 1 \) as \( \Delta^{(n)} \to 0 \). Because the summands of Lemma
1(a)(i) are uniformly bounded and because \( \Sigma \) is well-behaved, we have
\( Z_n \sim h(0, 1) \) and \( B^n \to 1 \) as \( n \to \infty \) and \( \Delta^{(n)} \to 0 \), and the proof for
the stratified analysis is complete. The proof for the collapsed analysis under the specified models is similar.

Expressions for $\omega^2$ in terms of contingency table parameters. It follows from (A8) that for the unmatched design we have

$$\omega_S^2 \text{ (unmatched)} =$$

$$n[\sum (d_i - c_i)a_i \hat{f}_i]^2 / \sum [d_i(1-a_i) + c_i a_i][(1-d_i)(1-a_i) + (1-c_i)a_i]a_i \hat{f}_i$$

where

$$a_i = \frac{\bar{r}_i}{\hat{r}_i + \bar{r}_i},$$

and

$$\omega_C^2 \text{ (unmatched)} = \frac{2n(\hat{d} - \bar{c})^2}{[2(\hat{d} + \bar{c}) - (\hat{d} + \bar{c})^2]}.$$  (A19)

For the matched design we have from (A9)

$$\omega_C^2 \text{ (matched)} = \frac{2n(\hat{d} - \bar{c})^2}{[2(\hat{d} + \bar{c}) - \sum (d_i + c_i)^2 \hat{f}_i]}$$  (A20)

and

$$\omega_C^2 \text{ (matched)} = \frac{2n(\hat{d} - \bar{c})^2}{[2(\hat{d} + \bar{c}) - (\hat{d} + \bar{c})^2]}.$$  (A21)

If (8) and (9) hold, so that $X$ is related to neither $D$ nor $F$, then the four $\omega^2$ values (A18) through (A21) are equal.

Expansions of $\omega^2$ about $H_0$. It is convenient to let

$$\xi_i = \frac{\delta_i}{1 - p_i}.$$  (A22)

The following expansions about $\Delta = 0$ are obtained in a straightforward manner from (A4) and (A18) through (A21). For either Model F or Model D we have
\[
\omega^2_S \text{ (unmatched) } = \\
\frac{1}{2} n \Delta^2 \left\{ \sum q_i (1-q_i) r_i \right\} \left\{ \sum \xi_i f(\pi_i) r_i \right\}^2 / \sum f(\pi_i) r_i + o(\Delta^2)
\]

where \( f \) is defined by (31), and

\[
\omega^2_C \text{ (unmatched) } = \\
\frac{1}{2} n \Delta^2 q(1-q) \left\{ \sum \xi_i \pi_i (1 - p \pi_i) \right\}^2 / (1-p)^2 + o(\Delta^2).
\]

For either Model F or Model D, we have

\[
\omega^2_S \text{ (matched) } = \frac{1}{2} n \Delta^2 \left\{ \sum q_i (1-q_i) r_i \right\} \left\{ \sum \xi_i \pi_i r_i \right\}^2 + o(\Delta^2); \quad \text{(A25)}
\]

and for Model D only, we have

\[
\omega^2_C \text{ (matched) } = \frac{1}{2} n \Delta^2 q(1-q) \left\{ \sum \xi_i \pi_i r_i \right\}^2 + o(\Delta^2). \quad \text{(A26)}
\]

5.2. Proofs for Section 3.2

Effect of matching on the odds ratio. In Model F, \( p_i = p \) and \( \Delta_i = \Delta \) for all \( i \). From (A4), we have for Model F

\[
(\hat{\theta} - \tilde{\theta}) = \sum (\hat{r}_i - \tilde{r}_i) c_i \\
= \frac{\Delta [1 - p(1+\Delta)]}{(1 + \Delta q)(1 - p(1 + \Delta q))} \left\{ \sum q_i g(q_i) r_i - q \sum g(q_i) r_i \right\} \quad \text{(A27)}
\]

where \( g(x) = x/\left[1 - p(1+\Delta x)\right] \). Since \( g(*) \) is an increasing function, the "covariance" in \{ \} in (A27) must be nonnegative, so that \( (\hat{\theta} - \tilde{\theta}) \)
has the same sign as \( \Delta \). In addition, it is easily seen from (A2) and
(A4) that \( (\hat{\alpha} - \tilde{\alpha}) \) and \( (\hat{\alpha} - \tilde{\alpha}) \) have the same sign as \( \Delta \). It follows that
\( (\hat{\alpha} - \tilde{\alpha}) \) has the same sign as \( \Delta \) but smaller absolute value. Now in
general we have from (A8) and (A9), respectively
\[
\theta_c = \hat{\alpha}(1-\bar{c})/\bar{\alpha}(1-\hat{\alpha})
\]  \hspace{1cm} (A28)

and

\[
\theta' = \hat{\alpha}(1-\bar{c})/\bar{\alpha}(1-\hat{\alpha}).
\]  \hspace{1cm} (A29)

In Model F, \(\theta = \theta_c\); (A28) and (A29) together with the preceding statement imply that \((\theta' - 1)\) has the same sign as \((\theta - 1)\) but smaller absolute value.

Asymptotic distributions of \(W^2\) statistics under \(H_0\). The asymptotic null distributions for \(W^2_S\) and for \(W^2_C\) (unmatched) follow from Theorem 1.

From (A20) and (A21), we have for the matched design

\[
\frac{\omega^2_C}{\omega^2_S} = 1 - \frac{\sum [(d_i + c_i) - (\hat{\alpha} + \bar{c})]^2 \hat{f}_i}{2(\hat{\alpha} + \bar{c}) - (\hat{\alpha} + \bar{c})^2}.
\]  \hspace{1cm} (A30)

From the definitions of \(W^2\) and \(\omega^2\) it is evident that \(W^2_C/W^2_S\) for the matched design is given by an expression like (A30) in which the probabilities \(d_i, c_i\) and \(\hat{f}_i\) have been replaced by the corresponding observed relative frequencies; it is clear, then, from (A30) that \(W^2_C\) is strictly less than \(W^2_S\) except in the very improbable event that the empirical counterpart of \((d_i + c_i)\) does not depend upon \(i\).

Asymptotic relative efficiencies of valid \(W^2\) tests. Under Model F, we have from (30), (31), (A3), (A12) and (A22) that, for all \(i, \pi_i = 1, f(\pi_i) = 1,\) and \(\xi_i = 1/(1-p).\) With these substitutions the ARE's (21) and (22) are immediate from (A23), (A24) and (A25).

5.3. Computations for Section 3.4.

Extremal distributions. The problem of maximizing \(v\) among exposure distributions bounded by fixed \(q^{(1)}\) and \(q^{(2)}\) can be approached in two stages. First, fix the mean \(q\) and maximize the numerator of (23); it is
well-known that such a variance maximum is achieved by a distribution on
two strata with \( q_1 = q^{(1)} \) and \( q_2 = q^{(2)} \). Second, among all such dis-
tributions, maximize the expression (23) with respect to \( r_1 \). It is
straightforward to show that the maximum \( v^* \) is achieved by

\[
    r_1^* = \sqrt{q^{(2)} (1 - q^{(2)})} / \{ \sqrt{q^{(1)} (1 - q^{(1)})} + \sqrt{q^{(2)} (1 - q^{(2)})} \};
\]

the extremal distributions used in Tables 3 and 4 are based on this
expression for \( r_1^* \).

Relative incidence. From (A1), (A3) and (A4), the relative incidence
(24) is given exactly in Model F by \( (1 + \Delta q_2)/(1 + \Delta q_1) \), or by
\( (1 + \theta q_2)/(1 + \theta q_1) \) when \( p \to 0 \). The latter expression was used in
calculating the fourth column of Table 4.

Collapsing in matched design. Substituting into (A29) the expres-
sions obtained from (A2) and (A4) by letting \( p \to 0 \) and \( \Delta \to 0 \) gives
the relation

\[
    \theta' = \theta[1 + (\theta-1) q(1-v)]/[1 + (\theta-1) [q(1-v) + v]],
\]

which was used in calculating the fifth column of Table 4. It is clear
from Lemma 1(a)(i) and from the definitions of the \( W^2 \) statistics that
the stochastic limit of \( W^2_C/W^2_S \) as \( n \to \infty \) in the matched design is
\( \omega^2_C/\omega^2_S \). The sixth column of Table 4 was computed from the expression
obtained from (A30) upon letting \( p \to 0 \) and \( \Delta \to 0 \) in (A2) and (A4).

5.4. Proofs for Sections 4.2 and 4.3

Effect of matching on the odds ratio. In general, the relationship
between the odds ratio \( \theta_1 \) and the incidence ratio \( \Delta_1 \) is given by

\[
    \Delta_1 = (1-p_1)(\theta_1-1)/[1 + p_1(\theta_1-1)] . \quad (A31)
\]
For Model D, assuming that \( \theta_1 = \theta \) and using (A2), (A4), (A28) and (A31), we have

\[
\theta_C = \theta \ h[\sum k(p_i)r_i] , \tag{A32}
\]

where

\[
h(x) = (1-p)x/p(1-\theta x)
\]

and

\[
k(x) = x/[1 + (\theta-1)x] .
\]

Since \( h(\cdot) \) is an increasing function with \( h[k(p)] = 1 \) and \( (\theta-1)k(\cdot) \) is concave, it follows from (A32) and Jensen's inequality that \( \theta_C \leq \theta \) when \( \theta \geq 1 \), and \( \theta_C \geq \theta \) when \( \theta \leq 1 \), so that \( (\theta_C - \theta) \) always has the same sign as \( (\theta-1) \) but smaller absolute value. The expansion (27) can be obtained from (A32). The expansion (29) can be obtained similarly upon using (A31) in (A4) and (A29).

Asymptotic relative efficiencies of \( \hat{\omega} \) tests. The relation (36) follows upon incorporating the definition of Model D into (A25) and (A26); note that the odds ratio may be stratum-dependent. Since from (A31)

\[
\Delta \xi_i = (\theta_i - \theta) + o(\theta_i - \theta) , \tag{A33}
\]

the assumption that \( \theta_1 = \theta \) for all \( i \) can be replaced in ARE calculations by the assumption that \( \xi_1 = \xi \) for all \( i \). Under this assumption, (32) and (34) follow from (A23) through (A26). The function \( f(\cdot) \) in (31) is strictly concave if \( p < 1 \); furthermore \( f(p) = 1 \), so that the inequality (33) follows from Jensen's inequality and is strict unless
\( \pi_i = 1 \) for all \( i \). The inequality (35) follows from (34) upon noting from (28) that \( 0 \leq \nu_p \leq 1 \).

The case of a rare disease. It can be seen from \( A4 \) that \( A10 \) holds when \( p \to 0 \) in Model D; the relation (37) then follows from \( A2 \), \( A19 \), \( A20 \) and \( A21 \).
References


Radhakrishna, S. (1965), Combination of results from several $2 \times 2$ contingency tables. Biometrics 21, 86-98.
