MATCHING IN CASE CONTROL STUDIES

Matching addresses issues of confounding in the DESIGN stage of a study as opposed to the analysis phase

A means of providing a more efficient stratified analysis rather than a direct means of preventing confounding, by increasing precision of estimates (reduction in SE)
Individual Matching

Controls are matched to cases on one or more attributes (i.e. age, gender, smoking status, etc). Each case/control pair then has identical values on the matching factors. Requires a more complex analysis than unmatched data—analytical complexity required to stratify on data not matched on. Each matched set defines it’s own stratum—can be viewed as a single “individual”

Frequency Matching

Match on cell instead of individual. Ex. Frequency matching on age and sex. If 20% of cases are 50-54 year old females, than controls are selected in such a way that 20% are also 50-54 years old and female. Does not require using a matched analysis, because you take a random sample of controls in that cell (50-54 year old females). But you have to wait until cases accumulate before controls are selected (unless you know distribution in advance of matching factors)
WHY MATCH?

1. To make control for confounding more efficient when sample size is small

   Without matching, control for confounding in the analysis will result in many strata with sparse data. By balancing the distribution across strata, the estimates of the OR will be more stable—smaller standard errors, and thus narrower confidence intervals.

2. Even if sample size is not small, if there are many confounders with many categories, data can be sparse in any given stratum. However, you may be able to use multivariate analysis instead.

3. If obtaining information from subjects is expensive. i.e running expensive lab tests on blood samples. Matching will insure control of confounding and will not lead to loss of information. If cost of matching is small compared to cost of expanding study size, matching is worthwhile.

4. Sometimes control of confounding only possible by matching—i.e., controlling for sibship.

   Alternatives to matching: frequency matching, use multivariate analyses to control confounding
DISADVANTAGES OF MATCHING

Time consuming

Can be expensive

Can’t always find an exact match

Matching can decrease study efficiency because the effort expanded in finding matched subjects could be spent on gathering information for a greater number of unmatched subjects.
Matching Criteria

Matching will increase efficiency only if the matching variables are associated with both the disease and the exposure.

The matching variable must also NOT be on the causal chain, i.e. if high fat diet is an exposure of interest, don’t match on high cholesterol or vice versa.

If matching is used, matched analysis must be used to take advantage of the matching. If matching was done appropriately, and matching is not taken into account in the analysis, the OR will be biased towards the null.

Matching allows you to assess the relationship to exposure and disease having already taken the confounding variable(s) into account, so you don’t need to adjust for these variables in the analysis.
AVOID OVERMATCHING

Term originally referred to “loss of validity in a case control study stemming from a control group that was so closely matched to the case group that the exposure distributions differed very little.” (Rothman, Modern Epidemiology)

Once you match on a factor, you can NOT analyze this factor in the analysis. You have to be assured that you do NOT want to assess the relationship of this factor to the disease. If you match on a variable that is associated with another variable of interest, you will have essentially matched on both of these variables.

Example: If you match on neighborhood (i.e census tract), you may also be matching on SES, if neighborhood is correlated with SES. So you would NOT be able to analyze SES as a potential “exposure” variable because you have made the cases and controls the same on this variable.

Ex. If female controls are matched to female cases, and vice-vers, you can NOT assess the role of gender on disease because you’ve made cases, controls similar on this variable. If you match on smoking status, you cannot assess the role of this factor in disease. In effect, you are matching on this factor to control confounding for this factor. But you are not concerned about assessing the impact of this factor on the disease.
MORE RECENT INTERPRETATIONS OF OVERMATCHING

CONCERNS WITH EFFICIENCY, NOT VALIDITY

Matching can result in LESS information if the expense of matching reduces the total number of study subjects.

Study efficiency:

Total information content of data/ total number of subjects

Cost efficiency:

Total information content of data/costs of study

CONTROLS SIMILAR TO CASES ON EXPOSURE WILL NOT CONTRIBUTE TO THE ANALYSIS—loss of efficiency

Unnecessary matching—

IF MATCHING FACTOR IS ASSOCIATED WITH DISEASE BUT NOT EXPOSURE, MATCHED ANALYSIS WILL BE STATISTICALLY LESS EFFICIENT

IF MATCHING FACTOR IS ASSOCIATED WITH EXPOSURE BUT NOT DISEASE, MUST USE MATCHED ANALYSIS, OTHERWISE ODDS RATIO WILL BE BIASED TOWARDS NULL IN UNMATCHED ANALYSIS. BUT VARIANCE OF ODDS RATIO WILL BE INCREASED COMPARED TO UNMATCHED ANALYSIS OF SAME SAMPLE SIZE.
SAMPLE SIZE IN MATCHED STUDY IS NUMBER OF MATCHED PAIRS (OR TRIPLETS, ETC).

SAMPLE SIZE IN UNMATCHED STUDY IS NUMBER OF CASES AND CONTROLS

DO NOT MATCH UNLESS MATCHING VARIABLE ASSOCIATED WITH DISEASE AND EXPOSURE. MORE EFFICIENT TO CONDUCT UNMATCHED STUDY OTHERWISE.
ANALYSIS METHODS FOR INDIVIDUALLY MATCHED CASE-CONTROL STUDIES

1. Rationale is to control at the *design stage* for potential confounders

2. Case-Control Pair; Dichotomous Exposure
   
   - Unit of Analysis is the matched case-control pair.

   - There are 4 possible outcomes with respect to the matched pair:
     
     ⇒ Case exposed; Control exposed
     ⇒ Case exposed; Control not exposed
     ⇒ Case not exposed; control exposed
     ⇒ Case not exposed; control not exposed
Usual Display of Matched Case-Control Data with Dichotomous Exposure

<table>
<thead>
<tr>
<th></th>
<th>Exposure</th>
<th>Present</th>
<th>Absent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cases</strong></td>
<td>Present</td>
<td>f</td>
<td>g</td>
<td>f+g</td>
</tr>
<tr>
<td></td>
<td>Absent</td>
<td>h</td>
<td>j</td>
<td>h+j</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>f+h</td>
<td>g+j</td>
<td>n</td>
<td></td>
</tr>
</tbody>
</table>
ODDS RATIO FOR MATCHED PAIR DICHOTOMOUS EXPOSURE CASE-CONTROL STUDIES

\[ OR = \frac{g}{h} \]
HEURISTIC DEMONSTRATION OF ODDS RATIO

• Suppose that for each of $I$ strata, the layout for a fourfold table is given by:

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>$a_i$</td>
<td>$b_i$</td>
<td>$m_{i1}$</td>
</tr>
<tr>
<td>Not Exposed</td>
<td>$c_i$</td>
<td>$d_i$</td>
<td>$m_{2i}$</td>
</tr>
<tr>
<td>Total</td>
<td>$n_{1i}$</td>
<td>$n_{2i}$</td>
<td>$N_i$</td>
</tr>
</tbody>
</table>

• From earlier discussion of stratified analysis, we recall that the Mantel-Haenszel odds ratio is given by:

$$OR_{MH} = \frac{\sum_{i=1}^{l} \frac{a_i d_i}{N_i}}{\sum_{i=1}^{l} \frac{b_i c_i}{N_i}}$$
- Consider a single matched pair. The four possible outcomes are shown below:

1.

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Not exposed</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

2.

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Not Exposed</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

3.

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Not exposed</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

4.

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Not exposed</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
Contribution to Odds Ratio from Tables of Type 1:

1.

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Not exposed</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Contribution to Numerator:

\[
\frac{a_id_i}{N_i} = \frac{1 \times 0}{2} = 0
\]

Contribution to Denominator:

\[
\frac{b_ic_i}{N_i} = \frac{1 \times 0}{2} = 0
\]
Contribution to Odds Ratio from Tables of Type 2:

2.

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Not exposed</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Contribution to Numerator:

\[
\frac{a_id_i}{N_i} = \frac{1 \times 1}{2} = \frac{1}{2}
\]

Contribution to Denominator:

\[
\frac{b_ic_i}{N_i} = \frac{0 \times 0}{2} = 0
\]
Contribution to Odds Ratio from Tables of Type 3:

3.

<table>
<thead>
<tr>
<th>Cases</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Not exposed</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Contribution to Numerator:

\[
\frac{a_i d_i}{N_i} = \frac{0 \times 0}{2} = 0
\]

Contribution to Denominator:

\[
\frac{b_i c_i}{N_i} = \frac{1 \times 1}{2} = \frac{1}{2}
\]
Contribution to Odds Ratio from Tables of Type 4:

4.

<table>
<thead>
<tr>
<th></th>
<th>Case</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Not exposed</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Contribution to Numerator:

\[
\frac{a_i d_i}{N_i} = \frac{0 \times 1}{2} = 0
\]

Contribution to Denominator:

\[
\frac{b_i c_i}{N_i} = \frac{0 \times 1}{2} = 0
\]
COMPUTATION OF ODDS RATIO FROM STRATIFIED ANALYSIS OF I MATCHED PAIRS

<table>
<thead>
<tr>
<th>Table Type</th>
<th>Number of Tables</th>
<th>Contribution to Numerator</th>
<th>Contribution to Denominator</th>
<th>Mantel-Haenszel Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>f</td>
<td>$f \times 0 = 0$</td>
<td>$f \times 0 = 0$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>g</td>
<td>$g \times 1/2 = g/2$</td>
<td>$g \times 0 = 0$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>h</td>
<td>$h \times 0 = 0$</td>
<td>$h \times 1/2 = h/2$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>j</td>
<td>$j \times 0 = 0$</td>
<td>$j \times 0 = 0$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>I</td>
<td>$g/2$</td>
<td>$h/2$</td>
<td>$g/h$</td>
</tr>
</tbody>
</table>

$$OR = \frac{\frac{g}{2}}{\frac{h}{2}} = \frac{g}{h}$$

Ratio of # of pairs of discordant exposure:

(#pairs with case exposed, controls not exposed, divided by #pairs with cases not exposed, controls exposed)

Case control pairs with same exposure NOT used

Matched pairs OR=$g/h$, or b/c
EXAMPLE OF MATCHED PAIR CASE-CONTROL ANALYSIS USING PAIRS MODULE

• Matched Case-Control Study of Association Between Use of Oral Conjugated Estrogens and Cervical Cancer (PEPI Manual Page 137)

<table>
<thead>
<tr>
<th>Controls</th>
<th>Estrogen Use</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present</td>
</tr>
<tr>
<td>Cases</td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>12</td>
</tr>
<tr>
<td>Absent</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
</tr>
</tbody>
</table>

OR=43/7=6.14
OUTPUT FROM PAIRS MODULE

PAIRS - Analysis of Paired Samples
Thursday, 3rd October 2002.

-------------------------------------------------------------------------
DATA
Number of "case = Yes, control = No" pairs: 43
Number of "case = No, control = Yes" pairs: 7

-------------------------------------------------------------------------
**IF NO CHI SQUARE TEST USE TWO TAILED P VALUE
 Doesn’t do unless enough pairs

One-tailed P = 0.000 [1.05E-07] Two-tailed P = 0.000 [2.10E-07]

Odds ratio = 6.14 or [reciprocal]: 0.16
90% conf. interval = 2.99 to 13.19 or [reciprocals]: 0.08 to 0.33
95% conf. interval = 2.66 to 14.93 or [reciprocals]: 0.07 to 0.38
99% conf. interval = 2.13 to 18.86 or [reciprocals]: 0.05 to 0.47

Low-bias indicator of O.R. = 5.38 or 0.16

-------------------------------------------------------------------------
SIGNIFICANCE TEST USED IS McNEMAR’S TEST
One Degree of Freedom Chi Square Test
p.442 Szklo

\[ \chi^2_{1df} = \frac{(|g - h| - 1)^2}{g + h} \]

FOR THIS EXAMPLE

\[ \chi^2_{MCNEMAR} = \frac{(|43 - 7| - 1)^2}{43 + 7} = 24.5 \]
APPROXIMATE CONFIDENCE INTERVALS FOR MATCHED Odds Ratios (p.442 Szklo)

\[ SE (\ln OR) = \sqrt{\frac{1}{b} + \frac{1}{c}} \]

95% CI(\ln OR) = \ln OR \pm (1.96 \times SE[\ln OR])

For 95% CI for OR take exponent

95% CI (OR) = \exp \left[ \ln (OR) \pm 1.96 \times \sqrt{\frac{1}{b} + \frac{1}{c}} \right]

FOR THIS EXAMPLE:

\[ \exp \left[ \ln(6.14) \pm 1.96 \times \sqrt{\frac{1}{43} + \frac{1}{7}} \right] = 2.76 - 13.64 \]

(PEPI gives slightly different CI (2.66-14.93) —uses exact methods)
NOTE: Cannot stratify on another variable that was not matching on—otherwise you will break the matching.

Example: If you matched only on age, but stratified on sex, age within gender would not necessarily be balanced on age. So if you want to control for both sex and age in a matched analysis, you must match on these factors in ADVANCE. Otherwise you can use logistic regression analyses, which retains the pairing, but allows for adjustment for other variables not matched on.
EXAMPLE FROM SCHLESSELMAN:

Table 7.19

Age and sex of three case control pairs, matched on age, but NOT matched on SEX

<table>
<thead>
<tr>
<th>Pair</th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M 20</td>
<td>F 20</td>
</tr>
<tr>
<td>2</td>
<td>M 30</td>
<td>F 30</td>
</tr>
<tr>
<td>3</td>
<td>F 40</td>
<td>M 40</td>
</tr>
</tbody>
</table>

Ages according to SEX—no longer matched on age

<table>
<thead>
<tr>
<th>Case</th>
<th>Male</th>
<th>Control</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>40</td>
<td>40</td>
<td>20</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>
INDIVIDUALLY MATCHED CASE CONTROL PAIRS WITH STRATIFICATION on ANOTHER MATCHING VARIABLE—Matched on Source of controls in addition to other matching variables

- Estrogen-Cervical Cancer Example Shown Earlier with “Augmented” Data (controls matched according to source of controls and other matching variables)

Stratum 1: Controls Selected from Hospitalized Patients

Stratum 2. Controls Selected from Population

<table>
<thead>
<tr>
<th>Cases</th>
<th>Estrogen Use</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Estrogen Use</td>
</tr>
<tr>
<td>Hospital</td>
<td>Present</td>
<td>12</td>
</tr>
<tr>
<td>Hospital</td>
<td>Absent</td>
<td>7</td>
</tr>
<tr>
<td>Hospital</td>
<td>Total</td>
<td>19</td>
</tr>
</tbody>
</table>

Population Controlled Study

<table>
<thead>
<tr>
<th>Cases</th>
<th>Estrogen Use</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Estrogen Use</td>
</tr>
<tr>
<td>Population</td>
<td>Present</td>
<td>9</td>
</tr>
<tr>
<td>Population</td>
<td>Absent</td>
<td>8</td>
</tr>
<tr>
<td>Population</td>
<td>Total</td>
<td>17</td>
</tr>
</tbody>
</table>
PAIRS — Analysis of Paired Samples
Thursday, 3rd October 2002.

DATA

STRATUM 1
- Number of "case = Yes, control = No" pairs: 43
- Number of "case = No, control = Yes" pairs: 7
- Number of "case = No, control = No" pairs: 121
- Number of "case = Yes, control = Yes" pairs: 12

STRATUM 2
- Number of "case = Yes, control = No" pairs: 37
- Number of "case = No, control = Yes" pairs: 8
- Number of "case = No, control = No" pairs: 104
- Number of "case = Yes, control = Yes" pairs: 9

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Odds ratio</th>
<th>Chi-square</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.14</td>
<td>28.82</td>
<td>1</td>
<td>P = 0.000 [7.95E-]</td>
</tr>
<tr>
<td>2</td>
<td>4.63</td>
<td>20.26</td>
<td>1</td>
<td>P = 0.000 [6.75E-06]</td>
</tr>
</tbody>
</table>

Chi-square for heterogeneity = 0.25 DF = 1 P = 0.614

POOLED DATA (Chi-square is continuity-corrected).

Odds ratio = 5.33 or [reciprocal]: 0.19
90% conf. interval = 3.26 to 8.84 or [reciprocals]: 0.11 to 0.31
95% conf. interval = 3.00 to 9.64 or [reciprocals]: 0.10 to 0.33
99% conf. interval = 2.54 to 11.40 or [reciprocals]: 0.09 to 0.39
Log-likelihood chi-sq. = 47.17 d.f. = 1 P = 0.000 [6.50E-12]
Pearson's chi-sq. = 43.12 d.f. = 1 P = 0.000 [5.16E-11]

IF THE STRATA ARE CLUSTERS OF RELATED OBSERVATIONS:
The above results require no modification (no positive correlation within clusters: Eliasziw-Donner rho = -0.02).
CAN MATCH MORE THAN ONE CONTROL PER CASE TO INCREASE PRECISION OF THE ODDS RATIO (DECREASE STANDARD ERROR)

GIVEN FIXED NUMBER OF CASES, PRECISION OF ODDS RATIO ESTIMATE DECLINES CONSIDERABLY FOR 5 OR MORE CONTROLS PER CASE

In other words, the increase in precision when matching 5 or more cases is minimal, and not worth the extra expense and resources required to conduct the matching.
INDIVIDUALLY MATCHED CASE-CONTROL STUDIES HAVING MORE THAN ONE CONTROL MATCHED TO EACH CASE

• Scenario: \( R \) controls are matched to each case

• Mantel Haenszel Chi-Squared Statistic is Shown Below:

\[
\chi^2 = \frac{\sum_{m=1}^{R} O_{1,m-1} - \frac{m}{R+1} (f_{1,m-1} + f_{0,m})}{\sum_{m=1}^{R} O_{1,m-1} + f_{0,m} \left( \frac{m(R+1-m)}{(R+1)^2} \right)}
\]

Where

\( f_{1,x} \) = The number of sets with the case exposed and \( x \) controls in the exposed category

\( f_{0,x} \) = The number of sets with the case not exposed and \( x \) controls in the exposed category

Mantel-Haenszel Odds Ratio:

\[
OR_{MH} = \frac{\sum_{m=1}^{R} (R + 1 - m) f_{1,m-1}}{\sum_{m=1}^{R} m f_{0,m}}
\]
The data below represent a case-control study of the relationship between history of induced abortions and tubal pregnancy. The 18 cases are women with tubal pregnancy; the controls are women not having tubal pregnancy. Each case has 4 matched controls; and history of induced abortions is designated with a ‘+’ indicating “yes” and a ‘-’ indicating ‘no’.

<table>
<thead>
<tr>
<th>Type</th>
<th>Case</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>0,0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1,1</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>1,0</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>0,0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0,1</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>1,0</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>1,0</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>0,0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1,2</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>1,1</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>1,2</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0,0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1,4</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>1,1</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>1,1</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>0,0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1,2</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
**SUPPOSE WE IGNORE THE MATCHING**

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Not Exposed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>12</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>Controls</td>
<td>16</td>
<td>56</td>
<td>72</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>72</td>
<td>90</td>
</tr>
</tbody>
</table>

CASECONT - Analysis of 2 X 2 Tables for Case-Control Studies

<table>
<thead>
<tr>
<th>Exposed</th>
<th>Not exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>12</td>
</tr>
<tr>
<td>Controls</td>
<td>16</td>
</tr>
</tbody>
</table>

ANALYSIS OF TABLE 1: Total cases = 18 Total controls = 72

Proportion of cases exposed = 0.667 Proportion of controls exposed = 0.222

Chi-square (1 DF) = 13.272 P = 0.001 [2.69E-04]

Continuity corrected chi-sq. (Yates) = 11.279 P = 0.001 [7.84E-04]

Upton's adjusted chi-square = 13.124 P = 0.001 [2.91E-04]

Odds ratio = 7.00 [Low-bias indicator of O.R. in the population = 5.65]

90% confidence interval = 2.38 to 21.34

95% confidence interval = 2.01 to 25.34

99% confidence interval (approximate) = 1.46 to 34.94

Adjusted O.R. (0.5 added in each cell) = 6.59

Yule's Q = 0.75 Phi = 0.38

Lambda (prediction of exposure status from "caseness") = 0.21

(prediction of "caseness" from exposure status) = 0.00
As with paired data, let us consider each matched set of 1 case and R controls as a single “stratum” which would yield the following fourfold table (NOTE CASE-CONTROL/EXPOSURE order reversed to be consistent with PEPI modules):

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Not Exposed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>y</td>
<td>1 - y</td>
<td>1</td>
</tr>
<tr>
<td>Controls</td>
<td>x</td>
<td>R - x</td>
<td>R</td>
</tr>
<tr>
<td>Total</td>
<td>x + y</td>
<td>R + 1 - (x + y)</td>
<td>R + 1</td>
</tr>
</tbody>
</table>

Note: $y = 0$ or 1 (Only one case, either exposed or not exposed)

- We can then compute the Mantel-Haenszel test and odds ratios as we do for a stratified analysis.

- The tables having $x = 0$ and $y = 0$; and $x = R$ and $y = 1$ are “non-informative” analogous to what we saw for the individually 1 to 1 matched case-control design.

- The Mantel-Haenszel test and odds ratio can then be calculated in the usual stratified analysis way (e.g., using the CASECONT module).
OUTPUT FROM CASECONT FOR 12 INFORMATIVE MATCHED SETS—EXCLUDES 1 SET WHERE ALL CASES and CONTROLS EXPOSED, AND 5 SETS WHERE ALL CASES AND CONTROLS NOT EXPOSED

CASECONT – Analysis of 2 X 2 Tables for Case-Control Studies
Saturday, 21st February 1998.

DATA

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Not exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Controls</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

ANALYSIS OF TABLE 1:    Total cases = 1    Total controls = 4

Proportion of cases exposed = 1.000   Proportion of controls exposed = 0.250

Chi-square (1 DF)                      =  1.875   P = 0.171
Continuity corrected chi-sq. (Yates) =  0.052   P = 0.819
Upton's adjusted chi-square          =  1.500   P = 0.221
** WARNING: 4 cells have an expected frequency of <5.

Odds ratio = infinity.  [Low-bias indicator of O.R. in the population = 1.50]
90% confidence interval (approximate) = 0.09 to infinity
95% confidence interval (approximate) = 0.06 to infinity
99% confidence interval (approximate) = 0.04 to infinity
Adjusted O.R. (0.5 added in each cell) = 7.00

Yule’s Q = 1.00        Phi = 0.61
Lambda (prediction of exposure status from "caseness") = 0.50
(prediction of "caseness" from exposure status) = 0.00
TABLE 2

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Not exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Controls</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

ANALYSIS OF TABLE 2:  
Total cases = 1  
Total controls = 4

Proportion of cases exposed = 1.000  
Proportion of controls exposed = 0.000

Chi-square (1 DF)  
= 5.000  
P = 0.025

Continuity corrected chi-sq. (Yates)  
= 0.703  
P = 0.402

Upton's adjusted chi-square  
= 4.000  
P = 0.046

** WARNING: 4 cells have an expected frequency of <5.

Odds ratio = infinity.  [Low-bias indicator of O.R. in the population = 4.00]

90% confidence interval (approximate)  
= 0.28 to infinity

95% confidence interval (approximate)  
= 0.20 to infinity

99% confidence interval (approximate)  
= 0.12 to infinity

Adjusted O.R. (0.5 added in each cell)  
= 27.00

Yule's Q = 1.00  
Phi = 1.00

Lambda (prediction of exposure status from "caseness")  
= 1.00

(precision of "caseness" from exposure status)  
= 1.00
### Table 3

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Not exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Controls</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

**Analysis of Table 3:**

- Total cases = 1
- Total controls = 4

Proportion of cases exposed = 0.000  Proportion of controls exposed = 0.250

- Chi-square (1 DF) = 0.313  $P = 0.576$
- Continuity corrected chi-square (Yates) = 0.000  $P = 1.000$
- Upton's adjusted chi-square = 0.250  $P = 0.617$

**WARNING:** 4 cells have an expected frequency of <5.

- Odds ratio = 0.00  [Low-bias indicator of O.R. in the population = 0.00]
- 90% confidence interval (approximate) = 0.00 to 249.77
- 95% confidence interval (approximate) = 0.00 to 418.72
- 99% confidence interval (approximate) = 0.00 to 1010.07
- Adjusted O.R. (0.5 added in each cell) = 0.78

- Yule's $Q = -1.00$  Phi = -0.25
- Lambda (prediction of exposure status from "caseness") = 0.00
- (prediction of "caseness" from exposure status) = 0.00
SIMILARLY, THE INFORMATIVE TABLES ARE ENTERED THROUGH TABLE 12

DATA

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Not exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Controls</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

ANALYSIS OF TABLE 12:  Total cases = 1  Total controls = 4

Proportion of cases exposed = 1.000  Proportion of controls exposed = 0.500

Chi-square (1 DF) = 0.833  P = 0.361
Continuity corrected chi-sq. (Yates) = 0.000  P = 1.000
Upton's adjusted chi-square = 0.667  P = 0.414
** WARNING: 4 cells have an expected frequency of <5.

Odds ratio = infinity.  [Low-bias indicator of O.R. in the population = 0.67]
90% confidence interval (approximate) = 0.03 to infinity
95% confidence interval (approximate) = 0.02 to infinity
99% confidence interval (approximate) = 0.01 to infinity
Adjusted O.R. (0.5 added in each cell) = 3.00

Yule's Q = 1.00  Phi = 0.41
Lambda (prediction of exposure status from "caseness") = 0.00
(prediction of "caseness" from exposure status) = 0.00
FINALLY, THE SUMMARY STRATIFIED ANALYSIS IS PERFORMED

SUMMARY ANALYSIS OF TABLES 1 to 12

Mantel-Haenszel chi-square (DF = 1)  = 16.000  P = 0.000  [ 6.33E-05 ]  
continuity corrected (DF = 1)  = 13.598  P = 0.000  [ 2.26E-04 ]  
NOTE: Due to small numbers, M-H test is not recommended.

Mantel-Haenszel odds ratio  = 33.00  
90% confidence interval  = 4.35 to 250.44  
95% confidence interval  = 2.95 to 369.18  
99% confidence interval  = 1.38 to 788.57  

Maximum-likelihood estimate of uniform odds ratio  = 78.80  
90% confidence interval (Cornfield-Gart)  = 6.61 to 5383.27  
95% confidence interval (Cornfield-Gart)  = 4.95 to 8529.62  
99% confidence interval (Cornfield-Gart)  = 2.93 to 19066.58  

Heterogeneity of O.R.'s: chi-sq (DF: 11)  = 9.11  P = 0.612  
Standardized rate ratio (standard: exposed group)  = 33.00  
Standardized rate ratio (standard: unexposed group)  not computed.
THE FORMULAS STATED EARLIER ARE “SHORTCUTS” TO AVOID HAVING TO ENTER EVERY MATCHED SET AS A SEPARATE TABLE

INSTEAD, TABULATE FREQUENCY OF EXPOSURE OUTCOMES AND ENTER INTO PEPI MODULE “MATCHED”

For this example:
DATA
Number of controls per case = 4
Case '+' and 0 control '+': 3 case-control sets
Case '+' and 1 control '+': 5 case-control sets
Case '+' and 2 controls '+': 3 case-control sets
Case '+' and 3 controls '+': 0 case-control set
Case '+' and 4 controls '+': 1 case-control set
Case '-' and 0 control '+': 5 case-control sets
Case '-' and 1 control '+': 1 case-control set
Case '-' and 2 controls '+': 0 case-control set
Case '-' and 3 controls '+': 0 case-control set
Case '-' and 4 controls '+': 0 case-control set

Mantel-Haenszel chi-square (1 DF)
without continuity correction    = 16.000    P = 0.000  [ 6.33E -05 ]

Walter's test
without continuity correction: z =  4.619    P = 0.000  [ 3.86E -06 ]
with continuity correction:    z =  4.547    P = 0.000  [ 5.45E -06 ]

Mantel-Haenszel estimate of odds ratio     = 32.97
Approximate 90% confidence interval     =  4.34 to 250.19
Approximate 95% confidence interval     =  2.95 to 368.83
Approximate 99% confidence interval     =  1.38 to 787.81

Maximum-likelihood estimate of odds ratio  = 22.57
Approximate 90% confidence interval     =  3.94 to 129.37
Approximate 95% confidence interval     =  2.82 to 180.74
Approximate 99% confidence interval     =  1.47 to 347.57

Low-bias estimator of odds ratio           = 16.50
FOR SITUATIONS WHERE THERE ARE A VARIABLE NUMBER OF CONTROLS MATCHED TO EACH CASE

• This is an important situation since intentions to match each case with $R$ controls are not often accomplished successfully.

• Mantel Haenszel Chi Squared Test:

$$\chi^2 = \frac{\sum_{R} \sum_{m=1}^{R} (f_{1,m-1} + f_{0,m}) - \frac{m}{R + 1} \cdot (f_{1,m-1} + f_{0,m})}{\sum_{R} \sum_{m=1}^{R} (f_{1,m-1} + f_{0,m}) \cdot \frac{m(R + 1 - m)}{(R + 1)^2}}$$
EXAMPLE OF INDIVIDUALLY MATCHED CASE-CONTROL STUDY WITH VARYING NUMBER OF CONTROLS PER CASE (FROM PEPI MANUAL PAGE 116)

- Cases are persons with myocardial infarctions (MI’s); exposure is coffee consumption at level of 6+ cups per day. Summary tables are shown below.

**Cases with 1 matched control:**

<table>
<thead>
<tr>
<th>Cases</th>
<th>Exposed Controls</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Exposed</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Not Exposed</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>11</td>
</tr>
</tbody>
</table>

**Cases with 2 Matched Controls:**

<table>
<thead>
<tr>
<th>Cases</th>
<th>Exposed Controls</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Exposed</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>Not Exposed</td>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>36</td>
<td>45</td>
</tr>
</tbody>
</table>
DATA

Total controls per case = 1
Case '+' and 0 control '+': 8 case-control sets
Case '+' and 1 control '+': 8 case-control sets
Case '-' and 0 control '+': 8 case-control sets
Case '-' and 1 control '+': 3 case-control sets

Total controls per case = 2
Case '+' and 0 control '+': 16 case-control sets
Case '+' and 1 control '+': 23 case-control sets
Case '+' and 2 controls '+': 4 case-control sets
Case '-' and 0 control '+': 20 case-control sets
Case '-' and 1 control '+': 22 case-control sets
Case '-' and 2 controls '+': 3 case-control sets

Mantel-Haenszel chi-square (1 DF)
without continuity correction = 7.792    P = 0.005  [ 5.25E-03 ]

Walter's test
without continuity correction: z = 2.714    P = 0.007  [ 6.64E-03 ]
with continuity correction:    z = 2.672    P = 0.008  [ 7.54E-03 ]

Mantel-Haenszel estimate of odds ratio = 2.06
Approximate 90% confidence interval = 1.35 to 3.14
Approximate 95% confidence interval = 1.25 to 3.41
Approximate 99% confidence interval = 1.06 to 3.99

Maximum-likelihood estimate of odds ratio = 1.98
Approximate 90% confidence interval = 1.32 to 2.99
Approximate 95% confidence interval = 1.22 to 3.23
Approximate 99% confidence interval = 1.04 to 3.77

Low-bias estimator of odds ratio = 1.96