META ANALYSIS

- *Meta-Analysis* is the process of using statistical methods to combine the results of different studies. In the biomedical sciences, it involves the systematic, organized, and structured evaluation of a problem of interest using information (commonly in the form of statistical tables or other data) from a number of independent studies of the problem.

- We are discussing it here because it is widely used in epidemiology and because it fits in nicely with our discussion of stratified analysis and the use of the PEPI module, *COMBINE*.

- For purposes of this course, we will limit our use of the term Meta Analysis to Quantitative Analysis of Published Data
Other Types of Synthesizing Results of Several Studies include:

1. Qualitative Review of published data with no use of combined statistical analyses—traditional literature review

2. Pooled Analysis from different studies using INDIVIDUAL data

3. Planned Prospective Study with a standardized protocol for several different geographical locations—also uses individual data and adjusts for geography
Some major issues in meta analysis:

Meta-analysis has traditionally been used to synthesize results from randomized clinical trials, where small sample sizes have led to unstable estimates, but where the protocols can be standardized for each study.

The use of meta-analysis in observational studies is more controversial, as critics point out the mixing of “apples and oranges,” i.e. the protocols, data collection methods, definitions of disease, exposures etc, will vary from study to study and hence results should not be combined. (See Blettner et al, “Traditional reviews, meta-analysis, and pooled analysis in Epidemiology”, Int J. Epidem 1999:28:1-9)

Another criticism is the issue of publication bias, in that positive studies are more likely to be published in peer reviewed journals than negative studies, hence biasing any meta analysis towards finding a significant result.
To address some of these issues, proposals have been published with recommendations for the reporting of results from meta-analysis, including specifications for reporting of background, search strategy, methods, results, discussion, and conclusions. (See consensus statement by Stroup et al, “Meta Analysis of observational studies in epidemiology,” JAMA. 2000;283:2008-2012.)
OVERALL OBJECTIVES IN META ANALYSIS OF EPIDEMIOLOGICAL STUDIES

1. To synthesize the individual study specific estimates of association into a single combined estimate (if justified). This is especially important in exploring the causal nature of low level associations.

2. To gain insight into relationships between study specific estimates of association and study specific characteristics. This insight can be used in the design of further studies.
Four Basic Steps in Conducting Meta Analysis

1. Identify studies with relevant data
2. Define inclusion criteria for studies
3. Abstract data
4. Statistical Analysis
Sources of Data

Computerized databases (MEDLINE, etc), index terms used, bibliographies, government publications, conference proceedings, unpublished data, English only publications?, etc.

MOOSE Criteria for Reporting Results

Detailed information on search criteria including qualifications of searchers (i.e. librarians, investigators, etc), search software used, years included, non-electronic sources used, etc.

Why so detailed?

If all relevant studies not included, could influence results.
Eligibility Criteria for Studies

Should be determined in advance, to reduce investigator bias.

Criteria include but are not limited to:

Types of studies included (case control, cohort, etc)

Years of publication covered

Languages

Restrictions on sample size

Definition of disease, exposures

Confounders that must be measured

Dose response categories similar
Abstracting of data

What should be abstracted from articles?

Should at least include:

Type of study, source of cases/controls or cohort, measures of association, confidence intervals, number of observations, what confounders have been adjusted for, if any, how disease, exposure, confounders were measured, response rates.
Statistical Analysis


A primary goal of meta analysis is to examine heterogeneity of effect

If studies are heterogeneous, then you shouldn’t combine them

If studies are NOT heterogeneous, than you can use a summary measure of effect
“The situations in which meta-analysis of many small and consistent studies yields a strong and credible conclusion based on estimation of a single summary measure of effect are the exception, not the rule. Calculation of a single summary estimate of effect size does not resolve the discrepancies among study results when these discrepancies are due to bias, confounding, or to differences in the selection criteria for subject, treatment or follow up. “

“Exploration of heterogeneity as a formal goal of meta analysis has replaced the simplistic use of meta-analysis to derive a single summary estimate of effect size. “

Petitti, D, “Meta analysis, Decision Analysis, and Cost-Effectiveness Analysis,” p. 214
Clinical vs. Statistical Heterogeneity—Pettiti, p. 214-215

Clinical heterogeneity (for clinical or observational studies):

Differences in the characteristics of studies, such as study design, loss to follow-up, response rates, differences in study subjects, intervention, treatment.

Statistical heterogeneity: “Incompatibility in the quantitative results of different studies.”

Statistical tests are used to detect the presence of statistical heterogeneity.

Clinical heterogeneity needs to be examined without the use of statistics.

Both types of heterogeneity need to be assessed.
Two types of models are used to produce summary effect measures

Random Effects Model

Fixed Effect Model
Random Effects vs. Fixed Effects Models

**Fixed effect model**

Inference is based on the studies actually done

The variance component of the summary effect is only composed of terms for the within study variance of each study.

**Random Effect Model**

Inference is based on the assumption that studies used in the analysis are a random sample of a hypothetical population of studies.

Variance component includes a between study component as well as a within study component. Hence, the random effects model will usually generate a confidence interval as wide or wider than that using the fixed effect model. Results from a random effects model will usually be more conservative (there are exceptions).
Random effects assumption—asks the question:

“Will the treatment produce benefit ‘on average’?”  
(for clinical studies)

or,

“Does the exposure produce an effect (positively or negatively) ‘on average’?”  
(observational studies)
**Fixed-effects** assumption asks the question,

“Did the treatment produce benefit on average in the studies at hand?”  (clinical studies)

or,

“Did the exposure have an effect (positively or negatively) on average, in the studies at hand?”  (observational studies)

There is NO agreement among statisticians as to which model is preferable. Some prefer the fixed effects model because the random effects model answers a question that is “abstruse and uninteresting.”

Others believe the question answered by the fixed effect model is not as interesting as that addressed by the random effects model.
HETEROGENEITY and Results using Fixed vs. Random Effect Models

If there is a lot of heterogeneity among studies, the between study variance will dominate the weights, and results can vary widely between fixed effects and random effects models. Fixed effect models weigh studies according to sample size, while random effects studies in this situation will weight the studies equally.

However, differences in the results of meta analysis using fixed and random effects models only occur when statistical heterogeneity exists. If no heterogeneity exists, than the choice of models is irrelevant because the results are identical (Greenland and Salvan, 1990, Thompson and Pocock, 1991).

These authors argue that if heterogeneity exists, then a summary measure should not be calculated.

One should instead examine the reasons for the heterogeneity.
FORMULAS FOR CALCULATING STATISTICAL TESTS OF HOMOGENEITY/HETEROGENEITY

Tests the hypothesis that the effect size is equal in all of the studies.

General form: Sum of the weighted difference between the summary effect measure (odds ratios, relative risks, etc), and measure of effect from each study. Follows chi square distribution, but is usually called “Q.” Degrees of freedom are one less than the total number of studies included.

Mantel Haenszel, Peto and General Variance Based formulas are three ways used in calculating Q.

For example:

Mantel Haenszel method:

\[ Q = \text{sum} \{ \text{weight}_i \times [\ln \text{OR}_{mh} - \ln \text{OR}_i]^2 \} \]

Weight = 1/variance\_i
Variance = ni / (bi x ci)
Results using different models will vary only if there is statistical heterogeneity. Using a random effects model does not take into account heterogeneity. Still should not calculate a summary measure of effect if heterogeneity exists.
LIMITATIONS OF STATISTICAL TESTS OF HETEROGENEITY

Size of the statistic is dependent on study weights

Study weights are a function of study size

When studies are small or total number of studies is small, statistic will in general be small

Statistical tests of heterogeneity have low power

If no statistical heterogeneity exists, clinical heterogeneity still needs to be examined.
Magnitude of study statistic for heterogeneity also depends on the magnitude of the deviation from study results from the summary estimate of effect size.

A single outlier can make a very large contribution to study statistic.

If all of the effect sizes are positive, than there could still be significant heterogeneity if there is a single outlier.
Example from Pettiti, p. 217

Meta analysis of six case control studies of family history as a risk factor for ovarian cancer

<table>
<thead>
<tr>
<th>Study</th>
<th>Odds Ratio</th>
<th>95% C.I.</th>
<th>Weight of Study to Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.3</td>
<td>0.5-173.1</td>
<td>.4</td>
</tr>
<tr>
<td>2</td>
<td>18.2</td>
<td>4.9-69.0</td>
<td>2.2</td>
</tr>
<tr>
<td>3</td>
<td>11.3</td>
<td>0.6-211.3</td>
<td>.5</td>
</tr>
<tr>
<td>4</td>
<td>3.6</td>
<td>1.8-7.2</td>
<td>8.0</td>
</tr>
<tr>
<td>5</td>
<td>3.3</td>
<td>1.1-9.4</td>
<td>3.2</td>
</tr>
<tr>
<td>6</td>
<td>1.9</td>
<td>1.1-3.6</td>
<td>12.9</td>
</tr>
</tbody>
</table>

Results based on fixed effects model:

*All studies included:*

Summary Odds ratio: 3.1 95% CI 2.1-4.5

Q=11.5

Df =5, p value homogeneity =.04

*Excluding study 2*

Summary odds ratio: 2.7, 95% CI 1.8-3.9

Q=4.0

Df =4, p value homogeneity=.41

With Study 2 included, would conclude there is statistical heterogeneity, even though all the odds ratios are elevated, because the value of the odds ratio for study 2 deviates widely from the summary odds ratio.

The heterogeneity results are counterintuitive, because all the results are elevated (but not all are significant!)
PETTITI ASSERTS THAT OUTLIERS SHOULD NOT BE EXCLUDED SOLELY ON THE BASIS OF STATISTICAL TESTS!

SIMILARLY, individual values in a case control study or clinical trial would not be discarded post hoc on the basis of statistical tests.

It should be determined in the design stage which studies should be included, not based on statistical tests.
Statistical methods for Meta Analysis to estimate summary effect measures (Pettiti, p. 101)

<table>
<thead>
<tr>
<th>Model Assumption</th>
<th>Methods</th>
<th>Effect Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Effects</td>
<td>Mantel -Haenszel</td>
<td>Ratios (Odds ratios, rate ratios, risk ratio)</td>
</tr>
<tr>
<td></td>
<td>Peto</td>
<td>Ratio (approximates Odds ratio)</td>
</tr>
<tr>
<td></td>
<td>General</td>
<td>Ratio (all types)</td>
</tr>
<tr>
<td></td>
<td>Variance Based</td>
<td>and rate difference</td>
</tr>
<tr>
<td>Random Effects</td>
<td>DerSimonian-Laird</td>
<td>Ratio (all types)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>and rate difference</td>
</tr>
</tbody>
</table>

None of the fixed effect methods incorporate confounding. Hence they are not very useful if the association is confounded by any other variables. They also require two by two tables from each study to calculate the summary effect.
GENERAL VARIANCE BASED METHODS THAT USE CONFIDENCE INTERVALS

Another method was developed using confidence intervals, where the effect measures only require information on the study’s estimate of relative risk (odds ratios) and its’ confidence intervals (Prentice and Thomas 1986) and Greenland (1987)

This method is used in PEPI COMBINE module
SUMMARY MEASURE USING CONFIDENCE INTERVALS

Summary measure: \[ \ln RR_s = \frac{\sum (w_i \times \ln RR_i)}{\sum w_i} \]

Where \( w_i = \frac{1}{\text{Variance } RR_i} \)

RR can be relative risks or odds ratios

Variance \( RR_i = \left[ \frac{(\ln RR_i / RR_i)}{1.96} \right]^2 \)

\( RR_i \) = RR estimate in ith study
\( RR_l \) = lower bound of confidence interval for ith study
\( RR_u \) = upper bound of confidence interval for ith study

When confidence interval is symmetric, i.e.

\( (RR_i / RR_l) = (RR_u / RR_i) \),

Variance \( RR_i = \left[ \frac{(\ln RR_u / RR_i)}{1.96} \right]^2 \)

95% Confidence Interval

\[ \exp \left[ \ln (RR_s) \right] \pm 1.96 \times \sqrt{\text{variance}_s} \]

\( \text{variance}_s = 1 / \sum \text{weights}_i \)
\( RR_s = \text{Summary } RR \)
USE OF \textit{COMBINE} MODULE FOR META-ANALYSIS
### EXAMPLE: U.S. STUDIES OF ETS (Environmental Tobacco Smoke) IN THE WORKPLACE AND LUNG CANCER

<table>
<thead>
<tr>
<th>Study (Acronym Refers to Author)</th>
<th>Significance</th>
<th>Odds Ratio or Risk Ratio</th>
<th>L95%CI</th>
<th>U95%CI</th>
<th>Odds Ratio or Risk Ratio</th>
<th>L95%CI</th>
<th>U95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. BROW</td>
<td>NS</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. FONT</td>
<td>S</td>
<td>1.12</td>
<td>0.91</td>
<td>1.36</td>
<td>1.39</td>
<td>1.11</td>
<td>1.74</td>
</tr>
<tr>
<td>3. GARF</td>
<td>NS</td>
<td>0.93</td>
<td>0.55</td>
<td>1.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. JANE</td>
<td>NS</td>
<td>0.91</td>
<td>0.8</td>
<td>1.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. KABAT1F</td>
<td>NS</td>
<td>0.68</td>
<td>0.32</td>
<td>1.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. KABAT1M</td>
<td>S</td>
<td>3.27</td>
<td>1.01</td>
<td>10.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. KABAT2F</td>
<td>NS</td>
<td>1.15</td>
<td>0.62</td>
<td>2.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. KABAT2M</td>
<td>NS</td>
<td>1.02</td>
<td>0.5</td>
<td>2.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. SCHWARTZF</td>
<td>NS</td>
<td>1.35</td>
<td>0.89</td>
<td>2.04</td>
<td>1.5</td>
<td>0.99</td>
<td>2.26</td>
</tr>
<tr>
<td>10. SCHWARTZM</td>
<td>NS</td>
<td>1.35</td>
<td>0.74</td>
<td>2.45</td>
<td>1.5</td>
<td>0.75</td>
<td>3.01</td>
</tr>
<tr>
<td>11. STOCK</td>
<td>NS</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>12. WU</td>
<td>NS</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1.3</td>
<td>0.5</td>
<td>3.3</td>
</tr>
<tr>
<td>13. BUTLCoH</td>
<td>NS</td>
<td>1.72</td>
<td>0.33</td>
<td>9.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. CARDCoH</td>
<td>NS</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>
SOME NOTES ON THE ETS WORKPLACE STUDIES

- Studies 1-12 are case-control studies.
- Studies 13-14 are cohort studies.
- Note that some give no quantitative estimates of association; some give crude OR’s or RR’s only; some give confounder controlled OR’s or RR’s only; some give both. The two types of OR’s should not be combined with each other.
- In our example, we will combine the crude odds ratios for case-control studies 2-10.
DATA

<table>
<thead>
<tr>
<th>Ratio</th>
<th>Lower CL</th>
<th>Upper CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.12</td>
<td>0.91</td>
<td>1.36</td>
</tr>
<tr>
<td>0.93</td>
<td>0.55</td>
<td>1.55</td>
</tr>
<tr>
<td>0.91</td>
<td>0.80</td>
<td>1.04</td>
</tr>
<tr>
<td>0.68</td>
<td>0.32</td>
<td>1.47</td>
</tr>
<tr>
<td>3.27</td>
<td>1.01</td>
<td>10.62</td>
</tr>
<tr>
<td>1.15</td>
<td>0.62</td>
<td>2.13</td>
</tr>
<tr>
<td>1.02</td>
<td>0.50</td>
<td>2.09</td>
</tr>
<tr>
<td>1.35</td>
<td>0.89</td>
<td>2.04</td>
</tr>
<tr>
<td>1.35</td>
<td>0.74</td>
<td>2.45</td>
</tr>
</tbody>
</table>

Overall measure of association = 1.00
90% confidence interval = 0.92 to 1.09
95% confidence interval = 0.91 to 1.11
99% confidence interval = 0.88 to 1.14
Chi-square (association) = 0.00  DF = 1  P = 0.963
Chi-square (heterogeneity) = 11.34  DF = 8  P = 0.183

DerSimonian-Laird summary measure = 1.06
90% confidence interval = 0.93 to 1.20
95% confidence interval = 0.90 to 1.24
99% confidence interval = 0.86 to 1.30
Chi-square (DF: 1) = 0.49  P = 0.486
Now use COMBINE for the Adjusted Odds Ratios
DATA

<table>
<thead>
<tr>
<th>Ratio</th>
<th>Lower CL</th>
<th>Upper CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.79</td>
<td>0.61</td>
<td>1.03</td>
</tr>
<tr>
<td>1.39</td>
<td>1.11</td>
<td>1.74</td>
</tr>
<tr>
<td>1.5</td>
<td>0.99</td>
<td>2.26</td>
</tr>
<tr>
<td>1.5</td>
<td>0.75</td>
<td>3.01</td>
</tr>
<tr>
<td>1.3</td>
<td>0.5</td>
<td>3.3</td>
</tr>
</tbody>
</table>

---

ANALYSIS OF 5 VALUES

<< Fixed-effect model >>

Overall measure of association = 1.16
90% confidence interval = 1.02 to 1.32
95% confidence interval = 1.00 to 1.35
99% confidence interval = 0.95 to 1.42

Chi-square (association) = 3.85  DF = 1  P = 0.050
Chi-square (heterogeneity) = 12.82  DF = 4  P = 0.012

---

<< Random-effects model >>

DerSimonian-Laird summary measure = 1.21
90% confidence interval = 0.93 to 1.59
95% confidence interval = 0.88 to 1.68
99% confidence interval = 0.80 to 1.86

Chi-square (DF: 1) = 1.40  P = 0.237
MORE EXAMPLES OF HETEROGENEITY (CLINICAL OR STATISTICAL) IN META ANALYSIS

Effect of exposure may vary in subgroups

Effect of exposure may vary in types of studies (methodological differences)
ESTROGEN REPLACEMENT THERAPY AND CORONARY DISEASE

1991 Meta analysis (Stampfer and Colditz)

Statistical heterogeneity existed in 31 different studies

Summary estimates were then calculated separately by study design

Stampfer and Colditz reasoned that the cohort studies with internal controls, and the cross sectional studies had the least bias, and used these types of studies to conclude that estrogen replacement therapy decreased the risk of coronary disease!
SUBGROUP ANALYSIS OF LONG TERM USE OF HORMONE REPLACEMENT THERAPY AND RELATIVE RISK OF BREAST CANCER (p. 225 Pettiti)

Selected subgroups—  

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>RR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LT 60</td>
<td>1.3</td>
<td>1.1-1.5</td>
</tr>
<tr>
<td>GE 60</td>
<td>1.4</td>
<td>1.2-1.7</td>
</tr>
<tr>
<td>Family History</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.4</td>
<td>1.2-1.5</td>
</tr>
<tr>
<td>Yes</td>
<td>1.1</td>
<td>.7-1.6</td>
</tr>
<tr>
<td>Ethnic Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.3</td>
<td>1.1-1.4</td>
</tr>
<tr>
<td>Other</td>
<td>1.2</td>
<td>.7-2.2</td>
</tr>
<tr>
<td>Age at First birth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 25 yrs</td>
<td>1.4</td>
<td>1.0-1.9</td>
</tr>
<tr>
<td>&gt;=25 yrs</td>
<td>1.3</td>
<td>1.0-1.5</td>
</tr>
</tbody>
</table>

14 total subgroups -- Summary estimates of relative risks were similar in all subgroups
EVEN when Statistical Heterogeneity does not exist, it may be beneficial to examine subgroup differences in summary estimates of risk.