Introduction to Meta-Analysis

Alfred Rademaker, Ph.D.

Department of Preventive Medicine
Feinberg School of Medicine
Northwestern University

March 10, 2015
Outline

• Basics, Forest plot
• Specific examples (Craft, Jennings)
• Fixed and Random models
• Jennings paper
• CMA software
• Publication bias
• Power
Meta analysis

A quantitative statistical analysis of several separate but similar experiments or studies in order to test the pooled data for statistical significance
Each meta-analysis has

• Multiple studies
• >= 1 condition in each study
  - a single defined population (1 condition)
  - treatment/control in a randomized study (2 conditions)
  - exposed/unexposed in an observational study (2 conditions)
• An effect size and its standard error for each study
An effect size may be

- A single proportion or mean or correlation coefficient
- A difference in means, usually standardized by dividing by some kind of standard deviation
- A ratio of means
- An odds ratio
- A risk ratio (or relative risk)
- A risk difference
- A hazard ratio
- Anything you want, as long as you can calculate it and its standard error
# Impact of Statin Dose On Death and Myocardial Infarction

<table>
<thead>
<tr>
<th>Study Name</th>
<th>Risk Ratio</th>
<th>Relative Weight</th>
<th>Risk ratio and 95% confidence interval</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prove-it</td>
<td>0.84</td>
<td>13%</td>
<td></td>
<td>0.106</td>
</tr>
<tr>
<td>A to Z</td>
<td>0.86</td>
<td>19%</td>
<td></td>
<td>0.096</td>
</tr>
<tr>
<td>TNT</td>
<td>0.80</td>
<td>31%</td>
<td></td>
<td>0.0002</td>
</tr>
<tr>
<td>Ideal</td>
<td>0.89</td>
<td>37%</td>
<td></td>
<td>0.069</td>
</tr>
<tr>
<td>Summary</td>
<td>0.85</td>
<td>100%</td>
<td></td>
<td>0.000</td>
</tr>
</tbody>
</table>

Favours high dose

Favours std dose
Summarizing a meta-analysis using a forest plot

Title
Each line represents an individual study
Last line represents summary of all studies
Study specific information given on the left side
Scale on horizontal axis is effect size scale
Vertical line drawn at null hypothesis value of no difference
For difference effect size measures, null hypothesis is at 0
For ratio effect size measures, null hypothesis is at 1
Horizontal scale is divided into favoring one ‘treatment’ or favoring the other
Individual study effect sizes indicated by solid square
Size of square indicates study sample size
Lines are 95% confidence interval for individual study effect size
Individual study p-values are given on right side
Summary effect size is solid diamond at bottom
Width of diamond represents 95% confidence interval of summary effect
P-value given for summary effect
Craft LL, Vaniterson EH, Helenowski IB, Rademaker AW, Courneya KS.

Exercise Effects on Depressive Symptoms in Cancer Survivors: A Systematic Review and Meta-analysis.


PMID: 22068286; PMCID: PMC3253916.
Cancer survivors
Depression
Exercise

Does aerobic exercise for cancer survivors reduce depression?
15 randomized controlled trials (RCTs)
14 papers (one paper had a trial for women in treatment and a trial for women who completed treatment)

• Mix of cancer patients
• Outcome was a depression scale where higher numbers indicated more depression symptoms
• Aerobic exercise program versus usual care
<table>
<thead>
<tr>
<th>Model</th>
<th>Study name</th>
<th>Subgroup within study</th>
<th>Comparison type</th>
<th>Statistics for each study</th>
<th>Std diff in means and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>02</td>
<td>Deley</td>
<td>1 Breast</td>
<td>Aerobic-Control</td>
<td>-0.982 (0.261)</td>
<td></td>
</tr>
<tr>
<td>03</td>
<td>Badger</td>
<td>1 Breast</td>
<td>Aerobic-Control</td>
<td>0.757 (0.269)</td>
<td></td>
</tr>
<tr>
<td>04</td>
<td>Monga</td>
<td>2 Prostate</td>
<td>Aerobic-Control</td>
<td>-0.464 (0.443)</td>
<td></td>
</tr>
<tr>
<td>05A</td>
<td>Cadmus (impact)</td>
<td>1 Breast</td>
<td>Aerobic-Control</td>
<td>-0.078 (0.238)</td>
<td></td>
</tr>
<tr>
<td>05B</td>
<td>Cadmus (YES)</td>
<td>1 Breast</td>
<td>Nonuse-Exercise-Control</td>
<td>-0.153 (0.245)</td>
<td></td>
</tr>
<tr>
<td>06</td>
<td>Muttie</td>
<td>1 Breast</td>
<td>MinExe-Control</td>
<td>-0.244 (0.152)</td>
<td></td>
</tr>
<tr>
<td>07</td>
<td>Payne</td>
<td>1 Breast</td>
<td>Aerobic-Control</td>
<td>-0.383 (0.478)</td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>Pena</td>
<td>1 Breast</td>
<td>MinExe-Control</td>
<td>-0.581 (0.327)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Thoren</td>
<td>8 Mix of diagnoses</td>
<td>MinExe-Control</td>
<td>-0.187 (0.193)</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Cooreyes</td>
<td>3 Colorectal</td>
<td>Aerobic-Control</td>
<td>-0.049 (0.220)</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Cooreyes</td>
<td>7 Lymphomas</td>
<td>Aerobic-Control</td>
<td>-0.517 (0.188)</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Cooreyes</td>
<td>1 Breast</td>
<td>Aerobic-Control</td>
<td>-0.034 (0.185)</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Colomann</td>
<td>2 Prostate</td>
<td>MinExe-Control</td>
<td>-0.289 (0.257)</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Kolteau</td>
<td>1 Breast</td>
<td>Aerobic-Control</td>
<td>-1.319 (0.425)</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Dodd</td>
<td>8 Mix of diagnoses</td>
<td>Aerobic-Control</td>
<td>-0.263 (0.234)</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2. Forest plot of effect size.**
Overall results (Figure 2 in paper) indicated that exercise intervention was effective

- mean change (exercise) = mean FU – mean Baseline, same for control
- \( d = \text{standardized difference} \)
- \( d = \frac{\text{mean change (exercise)} - \text{mean change (control)}}{S_{\text{within}}}/2 \)
- \( \text{Var}(d) = \frac{n_1+n_2}{n_1n_2} + \frac{d^2}{2(n_1+n_2)} \)
- Lower is better, negative differences indicate a decrease in depression
- Mean=-0.22, 95% CI (-0.43 to -0.009), p=0.04
- (Hedges’ g is a bias corrected version of d, but d is used in paper)
In the next few slides, will talk about 3 quantities

individual person measurement within study: $y_{ij}$

individual study effect size: $Y_i$

summary effect size: $M$
Fixed effect model

Individual person measurement within study

\[ y_{ij} = \theta + \alpha_i + \varepsilon_{ij} \]

i indexes study
j indexes person within study
\(\alpha_i\) fixed effect for study specific mean
\(\varepsilon_{ij}\) random effect for person within study

mean (\(y_{ij}\)) = \(\theta\)

\[ \text{var} (y_{ij}) = \sigma^2 = \text{within study variance} \]
• Fixed effect model in general assumes that all the $\alpha_i$ are fixed values with zero variance
• Fixed effect model in meta-analysis assumes all the $\alpha_i$ are zero, i.e. $y_{ij} = \theta + \varepsilon_{ij}$
• In either case,
  $\text{mean (} y_{ij} \text{)} = \theta$
  $\text{var (} y_{ij} \text{)} = \sigma^2$

ALSO: Normality of $y_{ij}$ is assumed
individual study effect size: $Y_i$

$Y_i = f(y_{ij} \text{ for all observations in } i’\text{th sample})$

e.g.
$Y_i = \text{standardized mean difference } \text{d in a 2 sample design}$
$Y_i = \text{mean of a single sample in a 1 sample design}$
$Y_i = \log(\text{odds ratio}) \text{ in a 2 sample design}$
$Y_i = \text{proportion in a 1 sample design}$

variance of individual study effect size: $V_{Y_i}$
Meta analysis summary effect size: $M$

$M = \text{weighted mean of study specific effect sizes } Y_i$

$W_i = \text{weight assigned to study } i$
\[ M = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i} \]
For the fixed effect model, the weights that minimize the variance of $M$ are

$$W_i = 1 / (\text{within study variance})$$

$$= 1/V_{yi}$$
\[ M = \sum_{i=1}^{k} \left\{ \frac{W_i}{\sum_{i=1}^{k} W_i} \right\} Y_i \]

Weights do not sum to 1
Relative weights sum to 1

Variance of \( M = V_M = \frac{1}{\sum_{i=1}^{k} W_i} \)
Random effects model

\[ y_{ij} = \mu + \zeta_i + \epsilon_{ij} \]
- \( \zeta_i \): random effect for study specific mean
- \( \epsilon_{ij} \): random effect for person within study

Mean (\( y_{ij} \)) = \( \mu \)

Var (\( y_{ij} \)) = \( \tau^2 + \sigma^2 \)

= between study variance \( \tau^2 \) + within study variance \( \sigma^2 \)

Dersimonian and Laird (1977) applied the basic random effects model to meta-analysis
How do you know whether to use the fixed effect model or the random effects model?

• Determine the variability in study specific effect sizes $Y_i$

• Determine whether this variability is more than you would expect under the fixed effect model

• If so, determine the estimate $T^2$ of the true between study variability so you can modify $V_{yi}$, the variance of $Y_i$
• Determine the variability in study specific effect sizes $Y_i$

a) $k=$ number of studies,

b) You have $Y_i$ and $V_{yi}$, i=1,...,k, can get fixed weights $W_i = 1/V_{yi}$

a) Run a fixed effects analysis to get M

b) Calculate $Q = \sum_{i=1}^{k} W_i (Y_i - M)^2$

c) $Q$ is a measure of total variability in $Y_i$
• Determine whether this variability is more than you would expect under the fixed effect model

a) Calculate \( Q = \sum_{i=1}^{k} W_i (Y_i - M)^2 \)

b) No true difference in study specific effect sizes → \( Q \) has a chi square distribution with \( k-1 \) df

→ \( E(Q) = k-1 \)

c) Determine the probability of exceeding \( Q \) using a chi square table with \( k-1 \) df. If this \( p<0.05 \) (say), then \( Q \) is significantly greater than \( k-1 \) and the random effects model is indicated.
If so, determine the estimate $T^2$ of true between study variability so you can modify $V_{yi}$, the variance of $Y_i$

- Calculate $T^2 = \max\{0, (Q - (k-1))/C\}$
  
  (C is a normalizing constant that is a function of $W_i$)

- Assuming a random effects model,
  $V_{yi}^* = V_{yi} + T^2 = \text{within study variance} + \text{between study variance}$

- $W_i^* = 1/(V_{yi} + T^2)$

- These weights result in the weighted least squares estimator of $\mu$ (true summary effect size)
\[ M^* = \sum_{i=1}^{k} \left\{ \frac{W_i^*}{\sum_{i=1}^{k} W_i^*} \right\} Y_i \]

Variance of \( M^* = V_{M^*} = 1/ \sum_{i=1}^{k} W_i^* \)
Higgins’ $I^2$

$I^2$ = the proportion of the observed variability that reflects real differences in effect size

$I^2 = 100 \times (Q - df)/Q$

$0 < I^2 < 100\%$
<table>
<thead>
<tr>
<th></th>
<th>g</th>
<th>Q-value</th>
<th>df (Q)</th>
<th>p-value</th>
<th>$T^2$</th>
<th>$I^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>-0.22</td>
<td>39.49</td>
<td>14</td>
<td>&lt;0.001</td>
<td>0.105</td>
<td>64.5%</td>
</tr>
</tbody>
</table>

$C = 242.15$

$T^2 = (39.49 - 14)/242.15 = 0.105$

$I^2 = 100\times(39.49 - 14)/39.49 = 64.5\%$
Random effects model weights

• Are less than fixed weights
• Are not as dependent on study specific sample size, i.e. the effect of different study sample sizes is dampened using random weights
• Result in a variance of summary effect size that is greater than fixed variance, so that when $T^2$ is not zero, random confidence intervals will be wider than fixed confidence intervals.
Publication Bias

- Significant results tend to be published
- Usually leads to overestimating summary effect size
- Key tool in assessing publication bias is the funnel plot
- Vertical axis – standard error (top represents more precision, lower se)
- Horizontal axis – effect size, centered around mean of all studies
• Ideal distribution of funnel plot – all the points are equally spread around the overall mean
• Bias observed if small studies with effects in the direction opposite to the hypothesized direction are missing – lower right or lower left of funnel plot have no studies
• Straight lines in funnel plot are +/- 2se so that points outside the lines have p<.05 in Forest plot
CORRELATION TEST

• Rank correlation of the points on a funnel plot
• If significant, then there is publication bias
• Test has low power, so in declaring it non-significant, you may be committing a Type II error
• Craft paper: n=15, r=-0.11, p=0.55
Galbraith Plot

• For each study, have $d$ (standardized difference), $SE(d)$ and $Z = d/se(d)$
• Like a tipped funnel plot (except it has $z$ instead of $d$ on Y axis and $1/se$ instead of reverse scale on x axis)
• If you perform a regression analysis of $Z$ vs $1/SE$ and constrain it to go through the origin, then the slope will be $M$, the fixed summary effect size
\( z_i = \frac{y_i}{\sqrt{v_i}} \)

\( x_i = 1/\sqrt{v_i} \)
REGRESSION TEST

• Egger et al. BMJ 1997;315:629
• Perform an unweighted regression of Z on 1/SE, not constraining it to go through the origin.
• If INTERCEPT is significant, then there is significant publication bias
• Craft paper:
  intercept = -0.73, se = 1.45, p = 0.62
  no evidence of publication bias
Power considerations in meta-analysis

Key factors in meta-analysis power

- Number of studies
- Sample size per study
- Effect size to be detected
- Variability in effect size
- \( \alpha = \Pr(\text{Type I error}) \)
- \( \beta = \Pr(\text{Type II error}) \)
- Power = 1 - \( \beta \)
Basic question in power analysis

How many studies, each of what size, are needed to detect a specified effect size with error probabilities $\alpha$ and $\beta$ when the data have a certain level of variability?
Notation and simplifying assumptions

• k studies
• Each study has 2 independent groups (e.g. randomized), each group of size n
• Effect size is a continuous measure comparing groups (e.g. d=standardized difference)
• Study specific d has variance= $\frac{1}{n}(2+d^2/4)$
• There is some common value of d that you want to have 80% power to detect
Implications of simplifying assumptions

• Since the sample size is the same for each study, and since the same $d$ is assumed for each study, each study weight $W_i$ is the same
• This results in an unweighted model
• Weights do not matter and do not need to be taken into account in any sample size formulae
• (study specific weights and sample size ratios could be assumed, but this would complicate the sample size calculation substantially)
• M is the unweighted mean of the d’s over all studies, i.e. \( M = \frac{\text{sum of di}}{k} \), where the di are independent of each other because they come from different studies.

• The variance of M

\[
\frac{\text{variance}}{k^2} = \frac{k[(1/n) (2+d^2/4)]}{k^2} = \frac{1}{kn}(2+d^2/4)
\]

• Using standard sample size derivation

\[
k_n = \frac{[Z_{1-\alpha} - Z_{\beta}]^2(2+d^2/4)}{d^2}
\]
• Assuming a fixed effect model is the same as assuming there is no intra-study correlation so it does not matter what combination of n and k you use.

• As long as you get kn total, you will achieve the required power (under all the simplifying assumptions of course).

• A reasonable way to proceed? Assume you can get n (or an average of n) per group per study, how many studies k do you need?

\[ k = \frac{[Z_{(1-\alpha)} - Z_\beta]^2(2+d^2/4)}{nd^2} \]
Example
Craft, post-hoc power calculation (even though I usually do not recommend them)

n=37 (average of intervention and control sample sizes across studies)
d=0.2
Var(d)=(1/n)(2+d^2/4) = .054
Assume α = 0.05, Z_{1-α/2} = 1.96
β = 0.20, Z_β = -0.84
k = [Z_{1-α} - Z_β]^2(2+d^2/4)/ nd^2
k = [1.96 + .84]^2(2+(.2)^2/4)/37(.2)^2 = 10.6 or 11
k = 7.84 [ .054 ] / .04

Need 11 studies to meet assumptions
Power in the random effects model

- More complicated
- Need to assume a value for $T^2$, the true between study variance, which may be difficult to get
- This will introduce an intra-study correlation which will increase the sample size requirement

$$k = \left[ Z_{(1-\alpha)} - Z_{(\beta)} \right]^2 \left[ \frac{(2+ d^2/4)}{n} + T^2 \right] / d^2$$
Example

Craft, $T^2=0.105$

$$k = [Z_{(1-\alpha)} - Z_\beta]^2 \left[ \frac{(2+ d^2/4)/n + T^2}{d^2} \right]$$

$$k = [1.96 + .84]^2 \left[ \frac{(2+.2^2/4)/37 + .105}{.2^2} \right]$$

$$k = 7.84 \left[ .054 + .105 \right]/.04$$

Between

$k = 31.2$ or 32

Within

Need 32 studies to meet assumptions under random effects model.