5 Introduction to meta-analysis

Definition: A quantitative combination of information from different sources, pertaining to a common research question.

- “New”: one of the fashionable modern techniques.
- Exponential increase in usage in medicine especially.
- Required by Government (!) in some contexts, e.g. Pharmaceutical Benefits Advisory Scheme (PBAC).
- Not straightforward statistically but the idea is good.

5.1 Motivation

Example 1: IV streptokinase:

- 1973, *Lancet*: Australian multi-centre trial: odds ratio for mortality ($R_X$ vs. placebo) was 0.77; 95% CI: (0.44, 1.32). Not much made of this, although benefit of 20% in mortality would be important!

<table>
<thead>
<tr>
<th>$OR &lt; 1$</th>
<th>$OR &gt; 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P &lt; 0.05$</td>
<td>2</td>
</tr>
<tr>
<td>$P &gt; 0.05$</td>
<td>3</td>
</tr>
</tbody>
</table>

Meta-analysis:

$OR = 0.78; 95\% \text{ CI: (0.64, 0.95)}$
• Another meta-analysis: Yusuf, 1985, 33 trials: $OR = 0.76$; 95% CI: (0.71, 0.80)

• 2 very large trials (1986, 1988) each showed $OR \simeq 0.8$, very significant.

• Change in clinical practice only came after publication of these large trials.

5.1 Motivation
Combining information: hardly a new idea!

History:

• Legendre and Gauss (early 19th C) and the orbit of comets.

• Pearson (1904): vaccine for enteric fever.

• Clarke (1920): determination of the values of the atomic weights of the elements.

• Cochran (1937): treatment effects from agricultural experiments.

5.1 Motivation

• Beecher (1955): the placebo effect.

• Gart (1962): statistical control of confounding.

• Glass (1976): the term “meta-analysis” enters the literature.

• 1980s: A number of books, social science focus

• 1986: DerSimonian and Laird: random effects approach.

• Late ’80s and on: a burgeoning statistical interest in meta-analysis.

5.1 Motivation
Modern reasons for meta-analysis:

• Dissatisfaction with narrative reviews.
• Systematically organise existing knowledge.

• Improved precision: there may be effects worth finding that are small, hence need very large sample sizes; save on resources.

• Useful for policy-makers and regulatory bodies in medicine especially.

5.2 Fixed effect approach

Consider the simplest approach:

1. $k$ studies on the same topic, giving estimates of an effect.

2. It is inferentially efficient to combine them.

3. Estimates $y_i$ with corresponding standard errors.

4. Obtain combined estimate as a weighted average of the $y_i$s.

5. Quote combined estimate and its standard error, or a confidence interval, or hypothesis test.

5.2 Fixed effect approach

In modelling terms:

$Y_i = \text{estimator in study } i$

$\theta_i = \text{parameter in study } i$

$W_i = \text{estimate of } [\text{Var}(Y_i)]^{-1}$

Large sample normal approximations are almost always used, so an appropriate scale may be sought (e.g. log(OR) rather than OR).

5.2 Fixed effect approach

Simple approach assumes that $\theta_i \equiv \theta$.

$Y_i = \theta + e_i, \quad i = 1, 2, \ldots k$

Obtain combined estimator as

$$\hat{\theta}_F = \frac{\sum W_i Y_i}{\sum W_i}$$
with standard error given by \[ \sqrt{\sum W_i} \].

Why use the reciprocal of the variance as the weights?

5.2 Fixed effect approach

- Compare with other cases, notably the pooled estimator of \( \sigma^2 \), 2 sample t-test.
- Why \((n_i - 1)\)?
- General result: weighted averages have minimum variance when the weights are inversely proportional to the individual variances.

This simple “fixed effect” model has been very widely used.

5.2 Fixed effect approach

Can lead to seductively simple methods of combining information; for \( k \) standardized mortality ratio studies \( i = 1, 2, ..., k \), with

\[
\text{SMR}_i = \frac{O_i}{E_i}
\]

the meta-SMR is just given by

\[
\text{meta-SMR} = \frac{\sum O_i}{\sum E_i}
\]

5.2 Fixed effect approach

Example: Stampfer meta-analysis (figure 1)

- Each trial is represented by its estimate and 95% CI
- Some details of interest may be included \((n, date, here)\)
- The scale is sometimes linear, sometimes log
- The combined estimate and its 95% CI are plotted
- Null line is often shown explicitly
- (Note annoying visual feature: the more precise a CI is, the less visually prominent it is!)
The important issues in meta-analysis are not all (directly) statistical.

5.3 Publication bias

The available research findings may be a selective sample from the studies actually conducted ("file-drawer" problem).

Suggested solutions:

- scan references of original and review articles
- computer searches

5.3 Publication bias

- tap informal networks to find out who’s doing what
- registries of studies (e.g. RCTs)
- funding bodies, drug companies
- change procedure for submission of articles so that result-specific content is removed (Miettinen)

5.3 Publication bias

Actual results:

“The OR for stomach cancer, comparing those with microwave ovens in the home to those without, was 3.2; 95% CI: (1.4, 7.3). The crude data are shown in the Table; the estimate adjusted for age, socio-economic status and vegetarian habit was 2.7; 95% CI (1.2, 6.1). These data therefore provide evidence that if you have a microwave oven in your home, you’re more likely to die from stomach cancer.”

<table>
<thead>
<tr>
<th>Microwave</th>
<th>Controls</th>
<th>Cases</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>66</td>
<td>51</td>
<td>117</td>
</tr>
<tr>
<td>Yes</td>
<td>10</td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td>Total</td>
<td>76</td>
<td>76</td>
<td>142</td>
</tr>
</tbody>
</table>

5.3 Publication bias

Suggestion for how this should be submitted to a journal:
"The OR for stomach cancer, comparing those with microwave ovens in the home to those without, was . The crude data are shown in the Table; the estimate adjusted for age, socio-economic status and vegetarian habit was . These data therefore provide evidence that ."

<table>
<thead>
<tr>
<th>Microwave</th>
<th>Controls</th>
<th>Cases</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td></td>
<td>117</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>76</td>
<td>76</td>
<td>142</td>
</tr>
</tbody>
</table>

5.3 Publication bias

- Publication bias can exist for secondary outcomes, even among published trials:

**Example 2**: β-blockers in AMI ( = acute myocardial infarction = heart attack): secondary outcome: ventricular fibrillation:

A. Trials which highlighted the results:

<table>
<thead>
<tr>
<th>n</th>
<th>Drug</th>
<th>Control</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1395</td>
<td>2.4%</td>
<td>0.33</td>
</tr>
<tr>
<td>2</td>
<td>477</td>
<td>5.6%</td>
<td>0.20</td>
</tr>
<tr>
<td>3</td>
<td>735</td>
<td>4.9%</td>
<td>0.38</td>
</tr>
</tbody>
</table>

B. Other trials (24 of them!)

<table>
<thead>
<tr>
<th>n</th>
<th>Drug</th>
<th>Control</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>24729</td>
<td>2.3%</td>
<td>2.5%</td>
<td>0.92</td>
</tr>
</tbody>
</table>

- Many of these data obtained by direct correspondence.

- It appears that authors are attracted by the "excitement" of results away from the null, especially when they are not specifically looking for them.

5.3 Publication bias

**Example 3**: Sinonasal cancer and wood dust exposures

- International Agency for Research on Cancer (IARC) did a pooled analysis of 13 case-control studies versus a meta-analysis of the published results.
• Primary concern is adenocarcinoma (AC), doubt about squamous cell carcinoma (SCC).

<table>
<thead>
<tr>
<th>Type</th>
<th>IARC pooling</th>
<th>Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR 95% CI</td>
<td>OR 95% CI</td>
</tr>
<tr>
<td>AC</td>
<td>13.5 (9.0, 20.0)</td>
<td>21.2 (13.3, 33.9)</td>
</tr>
<tr>
<td>SCC</td>
<td>1.0 (0.8, 1.4)</td>
<td>1.6 (1.1, 2.3)</td>
</tr>
</tbody>
</table>

• For most outcomes, meta-analysis OR > pooled analysis OR, suggesting a bias in the way results are published.

5.3 Publication bias

Inclusion decisions, forced or conscious, are possibly the most important effect on the results of a meta-analysis.

Conscious choices:
• exclude some studies (e.g. non-randomized)
• assess quality:

5.3 Publication bias

• One approach is to select widely: a badly designed study does not necessarily mean a biased result (Glass).
• Simes and others have used quality ratings, by independent researchers, to supplement the weights in the weighted average.

5.3 Publication bias

• “Trim and fill” technique (Tweedie): see figures 3 and 4.
• Assumes that the suppression of results for whatever reason has been on the basis of low estimates;
• Symmetrise the plot, re-analyse.

Beyond fixed effect:
General model is $Y_i = \theta_i + e_i$, assumption about $\theta_i$ determines whether we have a fixed or random effects model.

• Fixed effect [FE]: we assume that $\theta_i = \theta$:

$$Y_i = \theta + e_i,$$

$\Rightarrow$ uncertainty about the combined estimate, $\hat{\theta}_F$, is solely due to random variation within the studies.

• $\theta_i = \theta$ is the FE aspect of the FE model, or “homogeneity”.

• Formal tests of homogeneity (e.g. in logistic regression, Mantel-Haenszel).

• Cochran’s $Q$ statistic for testing $H_0: \theta_i = \theta$:

$$Q = \sum_{i=1}^{k} W_i(Y_i - \hat{\theta}_F)^2;$$

$H_0: \theta_i = \theta \Rightarrow Q \overset{d}{\approx} \chi^2_{k-1}$

Some approaches used to deal with heterogeneity:

• Test for systematic differences between studies (e.g. different outcomes according to study design).

• Model outcomes in terms of study characteristics: “meta-regression”.

• Refine inclusion criteria; it is often found that as criteria are made more restrictive the heterogeneity decreases.

• See examples of exercise and hypertension, figures 6 and 7.
“Random effects” models:

- The random effects [RE] model assumes that $\theta_i$ is itself a random variable, with mean $\theta$ and variance $\tau^2$.

- RE case: model above can be re-written as

$$Y_i = \theta + \epsilon_i + e_i,$$

with $\mathbb{E}(\epsilon_i) = 0$ and $\text{Var}(\epsilon_i) = \tau^2$.

5.4 “Random effects” models:

- ⇒ two contributions to the uncertainty about $\hat{\theta}$: the variation between the studies ($\epsilon_i$s) and the variation within studies ($e_i$s).

- Estimation of $\tau^2$ is done by considering the variation among the $\hat{\theta}_i$s, so the effective sample size is $k$.

5.4 “Random effects” models:

- The best-known implementation: DerSimonian and Laird (1986).

- The FE assumption ($\theta_i = \theta$) can be restated as $\tau^2 = 0$. In general,

$$\mathbb{E}(Q) \approx k - 1 + \tau^2 \left( \sum W_i - \frac{\sum W_i^2}{\sum W_i} \right),$$

- (Recall that $Q \approx \chi^2_{k-1}$ if $\theta_i = \theta$.)
• This leads to a method of moments estimator of $\tau^2$:

$$\hat{\tau}^2 = \max \left( 0, \frac{Q - (k - 1)}{\sum W_i - \frac{\sum W_i^2}{\sum W_i}} \right)$$

and new weights:

$$W_i^* = \frac{1}{W_i + \hat{\tau}^2}$$

5.4 “Random effects” models:

• Obtain combined estimator $\hat{\theta}_R$ as

$$\hat{\theta}_R = \frac{\sum W_i^* Y_i}{\sum W_i^*}$$

with standard error given by $\left[ \sum W_i^* \right]^{-\frac{1}{2}}$

• RE 95% CI for $\theta$: $\hat{\theta}_R \pm 1.96 \times se(\hat{\theta}_R)$

5.4 “Random effects” models:

Consider three cases:

• If we estimate that $\tau^2 = 0$, which happens when $q < k - 1$ (sometimes): RE inference (estimate, CI) reduces to the FE inference.

• If $\hat{\tau}^2$ is very large relative to the $\frac{1}{W_i}$s:
  - all studies get approximately equal weight, regardless of the within-study standard error (hence, regardless of sample size);
  - RE estimate is approximately a simple average of the study estimates.

5.4 “Random effects” models:

• In between these extremes the effect is that smaller (larger) studies tend to get more (less) weight in the RE estimate than in the FE one.

The typical RE outcome (e.g. Stampfer meta-analysis, figure 2):
• RE point estimate $\simeq$ FE point estimate
• $\text{se}(\hat{\theta}_R) > \text{se}(\hat{\theta}_F)$

5.5 Magnesium example

Example 4: Magnesium in suspected AMI: Does intravenous magnesium reduce mortality in cases of suspected heart attacks?

Background: Environmental evidence, autopsies animal data biochemical evidence safety, simplicity

1991 overview of 7 randomized trials (Teo, BMJ):

<table>
<thead>
<tr>
<th>Trial</th>
<th>Magnesium (Deaths/No. patients)</th>
<th>Control (Deaths/No. patients)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/40</td>
<td>2/36</td>
</tr>
<tr>
<td>2</td>
<td>9/135</td>
<td>23/135</td>
</tr>
<tr>
<td>3</td>
<td>2/200</td>
<td>7/200</td>
</tr>
<tr>
<td>4</td>
<td>1/48</td>
<td>1/46</td>
</tr>
<tr>
<td>5</td>
<td>10/150</td>
<td>8/148</td>
</tr>
<tr>
<td>6</td>
<td>1/59</td>
<td>9/56</td>
</tr>
<tr>
<td>7</td>
<td>1/25</td>
<td>3/23</td>
</tr>
<tr>
<td>Overall</td>
<td>25/657</td>
<td>53/644</td>
</tr>
<tr>
<td></td>
<td>(3.8%)</td>
<td>(8.2%)</td>
</tr>
</tbody>
</table>

5.5 Magnesium example

Meta-analysis: see figure.

Subsequent to the meta-analysis, there were two (much) larger trials (Yusuf & Flather, BMJ, 1995):

• LIMIT-2, which had 2316 patients: $OR = 0.74$, 95% CI: 0.55 – 1.00.

• More dramatically, ISIS-4, with 58,050 patients: $OR = 1.06$, 95% CI: 1.00 – 1.12.

5.5 Magnesium example

Raises an important issue of the evidence of meta-analysis versus a “mega-trial”: which do we believe?

In several cases they have agreed (Chalmers & Lau, SMMR, 1993):
• Mixed beta-blocking drugs versus no blocking of sympathetic receptors in AMI: two meta-analyses showed the same size reduction in death rate as two large trials.

5.5 Magnesium example

• Phototherapy for neonatal hyperbilirubinemia: meta-analysis of small RCTs in 1988 showed it to be very effective long before a large trial showed the same thing.

5.5 Magnesium example

The magnesium story:
A number of suggestions have been made to account for the discrepancy; notably, publication bias (small studies that showed that magnesium didn’t work may not have been published).

5.5 Magnesium example

Statistical issues were essentially discounted:

“Finally, it should be kept in mind that, with hundreds of meta-analyses being performed, a few will produce misleading results by chance alone—although this is unlikely in the present case.”

Hmmmmmm . . .

5.5 Magnesium example

The meta-analysis above was a fixed-effect analysis. See figure 13 for results of applying other methods. Furthermore:
Brockwell and Gordon (2001): the DerSimonian & Laird CI has inadequate coverage:
• See figure 9; DL = DerSimonian & Laird
• Reason: \( \tau^2 \) is estimated using a “small” sample size \((k)\), with non-trivial uncertainty.
• Analogous to the use of $t$ rather than Normal when estimating $\sigma^2$. 