Practical Considerations in Meta-Analysis.

Data Analysis Discussion Group

1st April, 2005

- Introduction and an example
- Epidemiological considerations
- Basic statistical methods
- Advanced statistical methods
- Software
- Literature
Introduction

- A *meta-analysis* is a statistical analysis of a collection of studies:
  - studies are the primary units of analysis.
  - focus on contrasting and combining results from different studies.

- Part of a broader goal of *research synthesis*.
  - bring together current knowledge on a given topic.
  - narrative components.
  - quantitative components.

- Ideal:
  - access to *all* studies conducted on specific topic.
  - subject-specific information/data.

- Primary difficulty:
  - heavy reliance on published literature.
  - publication bias; authors as well as editors.
  - quality/extent of information in published articles.
• Single-dose ibuprofen for post-operative pain.
  • non-steriodal anti-inflammatory (NSAID) analgesic.
  • 46 placebo-controlled trials between 1977-96.

• Warn, Thompson and Spiegelhalter (2002)
  • Cochrane Review.

• Outcome:
  • at least 50% pain relief in 4-6 hours after administration.
  • consider ‘risk’ of experiencing pain relief.

• Exposure:
  • doses range from 50mg to 800mg.
  • concentrate of 31 trials with dose 400mg.

• Initial impression:
  • considerable evidence that ibuprofen improves pain relief.
  • heterogeneity in effect sizes and uncertainty.
  • samples sizes vary from 28 to 391; median 80
Ahlstrom1993
Arnold1990
Bakshi1994
Cooper1977
Cooper1982
Cooper1988a
Cooper1989
Forbes1984
Forbes1990
Forbes1991b
Forbes1992
Frame1989
Frieke1993
Gay1996
Heidrich1985
Hersch1993a
Hersch1993b
Jain1986
Jain1988
Laska1986
Lavernziana1996
McQuay1996a
Mehisch1990
Mehilsch1995
Pagnoni
Schachtel1989
Seymour1991i
Seymour1991ii
Seymour1996
Sunshine1983
Sunshine1987

Data Analysis Discussion Group, 1st April, 2005.
Quantitative goals

• Improve statistical precision/power:
  • combine information from many small studies.

• Estimation of an overall/average ‘effect’:
  • summarize knowledge via a single quantity.

• Description of between-study heterogeneity:
  • determines ability to draw overall conclusions.
  • random vs systematic variation.

Strategy

• Epidemiological considerations.

• Re-analysis of individual studies.

• Descriptive/graphical methods.

• Regression-based methods.
Epidemiological considerations

Specification of study variables

- Outcome, exposure, confounders, effect modifiers, etc.

- Critical in assessing eligibility criteria for inclusion into the meta-analysis:
  - varying definitions across studies.
  - understanding of mechanisms evolve over time.
  - new/better adjustment for confounders.

- Helps understand/qualify compromises:
  - less control than when conducting an individual study.

Study identification

- Search of the published literature (e.g. ISI Web of Science).

- Search of computerised databases (e.g. MEDLINE).

- Inquire among researchers in the topic area.
Heterogeneity

- Determines relevance of conclusions drawn.

- Variety of strategies for heterogeneity:
  - regression models.

- Petitti (2001) interesting review of how heterogeneity is often dealt with.
  - statistical tests often not reported properly, or even performed.
  - consequences of results of tests of heterogeneity unclear.

- Recommendations:
  - explanation of model choice.
  - conduct of several types of analyses.
  - consideration of relevance of a summary estimate.
  - exploration of reasons for heterogeneity, if present.

- Selective exclusion to make the results homogeneous is inappropriate:
  - ‘outlier’ studies may contain important information.

- Focus should be on understanding clinical sources of heterogeneity.
Publication bias

- Many steps in the publication process:
  - investigator and publisher-related.

- Sample size plays a role:
  - large studies tend to be well designed.
  - data may be of higher quality.
  - higher statistical power.

- Preferential treatment to ‘statistically significant’ results.
  - yield results away from the null
  - essentially meaningless without clinical significance or sample size.
Re-analysis of individual studies

Goal

- Obtain study-specific information that is of a common format.
- Ideally obtain raw data from each study.
  - transform these into a common format, merge and analyse in a pooled analysis.
- Generally rely on published reports.
  - extent of information may vary across articles.

Extraction of information

- Choice of effect measure:
  - mean, slope, risk difference, relative risk, odds ratio, etc.
  - depends on study design characteristics (e.g. case-control study).
- Extraction of estimates:
  - point estimates and uncertainty estimates.
  - latter important to understand study-specific contributions.
  - may be obtained by standard errors, CI, or p-value.
  - rounding in published tables suggests you may only be able to obtain crude bounds.
Specific problems with p-values:

- sampling distribution: Normal, t, $\chi^2$.
- highly significant: $< 0.001$?
- not significant: $> 0.05$?

Sample size:

- original size vs. sample size used in analysis.
- missing information in, say, adjustment variables.

 Adjustment for confounding

- Partial or no adjustment for important confounding variables:
  - data collection inadequate.
  - biological understanding evolves over time.

- Crude approach using external estimates of confounding:

$$\theta_u = \theta_a \times \gamma$$

- $\gamma$ multiplicative bias.
- may be able to estimate $\gamma$ from other studies.
- standard errors also require adjustment.
- sensitivity analysis may be best approach.

- Rothman and Greenland (1998), chapter 32.
Notes

- Ability to perform extraction may be severely hampered by inadequate information.

- Even crude adjustments require a minimum of information.
  - some studies may have to be excluded.

- One option is to contact authors for more detailed information.
Descriptive/graphical methods

Tabulation of individual studies

- Publication status.
- Study design:
  - RTC, cohort study, case-control study.
  - design characteristics, e.g. follow-up.
- Study sample size:
  - cases/non-cases in treatment/control groups.
- Systematic differences:
  - variations in outcome definition.
  - e.g. coronary death, coronary death or MI, MI
- Exposure information:
  - number exposure/unexposed.
  - exposure quantiles.
- Effect estimates and standard errors:
  - information relevant to re-analysis.
Meta-analysis of cohort studies of daily coffee consumption and myocardial infarction or coronary death (1968-1992).

- selected entries from Rothman and Greenland (1998), pg 659.

<table>
<thead>
<tr>
<th>Study</th>
<th>Event</th>
<th>Cases</th>
<th>% ≥ n cups/d</th>
<th>Slope</th>
<th>SE</th>
<th>RR for 5 cups/d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Klatsky et al.</td>
<td>M</td>
<td>464</td>
<td>22 ≥ 7</td>
<td>-44</td>
<td>29</td>
<td>0.80</td>
</tr>
<tr>
<td>(1973)*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dawber et al.</td>
<td>M</td>
<td>322</td>
<td>15 ≥ 5</td>
<td>-39</td>
<td>40</td>
<td>0.82</td>
</tr>
<tr>
<td>(1974)†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wilhelmsen et al.</td>
<td>M</td>
<td>60</td>
<td>50 ≥ 5</td>
<td>109</td>
<td>153</td>
<td>1.72</td>
</tr>
<tr>
<td>(1977)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heyden et al.</td>
<td>D</td>
<td>36</td>
<td>13 ≥ 5</td>
<td>-44</td>
<td>76</td>
<td>0.80</td>
</tr>
<tr>
<td>(1978)†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Murray et al.</td>
<td>D</td>
<td>721</td>
<td>31 ≥ 5</td>
<td>-4</td>
<td>19</td>
<td>0.98</td>
</tr>
<tr>
<td>(1981)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grobbee et al.</td>
<td>DM</td>
<td>221</td>
<td>20 ≥ 4</td>
<td>4</td>
<td>30</td>
<td>1.02</td>
</tr>
<tr>
<td>(1990)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Nested case-control study
† Corrected estimates: published plus external adjustment
L’Abbe plot

- Risk in treatment group vs. risk in control group.
- Each trial represented by a symbol of area proportional to its precision:
  - inverse variance of the log-odds ratio.

Single-dose 400mg ibuprofen vs post-operative pain
(31 studies)
**Useful additions:**

- plotting symbols proportional extent of information
- reference value (e.g. null hypothesis).
- overall summary estimate (see later).
- groupings according to study characteristics.
Funnel plot

- May be useful in detecting publication bias.
  - large studies provide 'better' estimates.
  - small studies subject to variation around 'truth'.

- Plot effect size versus study size:
  - sample size or standard error.
  - expect a funnel/pyramid shape.

Single-dose 400mg ibuprofen vs post-operative pain
(31 studies)
Useful additions:

- study-specific confidence intervals.
- overall summary/average effect (see later).
- ‘mirror’ effect (see below).

Question: under-representation of studies around the null?
Basic statistical methods

Estimation of a common effect

• Weighted average of study-specific estimates.

• Weighting by ‘amount of information’, $w_i$:
  - sample size.
  - inverse variance.

• Let $\hat{\theta}_i$ denote the study-specific effect and $\hat{\sigma}_i$ the associated standard error estimate.

• Let $\mu$ denote the overall common effect:

$$\hat{\mu} = \frac{\sum w_i \hat{\theta}_i}{\sum w_i} \quad \text{and} \quad \widehat{\text{var}}[\hat{\mu}] = \frac{1}{\sum w_i}$$

where $w_i = 1/\hat{\sigma}_i^2$.

• ibuprofen:

  Study-specific ORs: vary between 1.29 and 126.50
  Common OR: 7.73 95% CI (6.29, 9.51)

• calculations performed on the log-odds ratio scale and transformed.

• Relevance relies heavily on a homogeneity assumption.
Summary

Data Analysis Discussion Group, 1st April, 2005.
Between-study heterogeneity

- Study $i$, $i = 1, \ldots, k$

$$\theta_i = \mu + \epsilon_i \quad \text{where } V[\epsilon_i] = \tau^2.$$  

- $\tau^2$ is the *between-study variance*.
- interpretation of $\tau^2$ depends on choice of $\theta$.
- estimation in the context of a random effects model (see later).
- ibuprofen:

$$\theta \equiv \text{log-odds ratio} \Rightarrow \hat{\tau}^2 = 0.53$$

- moderate due to large within-study uncertainty.

- Cochrans’ $\chi^2$ test or $Q$-test:

  - $H_0: \tau^2 = 0$

$$Q_w = \sum_{i=1}^{k} w_i (\hat{\theta}_i - \hat{\mu})^2 \sim \chi^2_{k-1}$$

  where $w_i$ is a study-specific weight.

- ibuprofen:

$$Q_w = 73.04 \text{ on } 30 \text{ df (p-value < 0.0001).}$$

- *caution*: generally low power.
- easy to overinterpret a non-significant result!
• Higgins and Thompson (2002):
  • 3 new measures that build on the $\hat{\tau}^2$.
  • quantify the impact of heterogeneity.
  • invariant to:
    (i) number of studies, $k$:
    (ii) choice of effect measure, $\theta$.

**Publication bias**

• Funnel plot asymmetry tests:
  • $H_0$: symmetry.

• Rank correlation method:
  • Begg and Mazumdar (1994)
  • standardised treatment effects versus variance estimates.

• Regression-based method:
  • linear regression of standardised treatment effect on the inverse of the standard error.

• Ibuprofen: p-values 0.0014 and < 0.0001 respectively.

• **Caution**: generally low power:
  • especially when number of studies is small (say, 25).
  • difficulty when interpreting a non-significant result.
Regression-based methods

Fixed effects model

- Overview by Brockwell and Gordon (2001).

- Study $i$, $i = 1, \ldots, k$

  $$Y_i = \theta_i + e_i$$
  where $e_i \sim N(0, \sigma_i^2)$

- $Y_i$ study-specific point estimate:
  - e.g. observed log-odds ratio.

- $\sigma_i^2$ is the within-study variation:
  - estimated via squared study-specific standard error estimates.

- Overall effect assume $\theta_i \equiv \mu$, for all $i = 1, \ldots, k$:
  - assumption can be assessed using above methods.

- Estimate of $\mu$ equivalent to weighted average outlined above:

  $$\hat{\mu} = \frac{\sum w_i Y_i}{\sum w_i}$$
  where

  $$\text{var}[\hat{\mu}] = \frac{1}{\sum w_i}$$
  where $w_i = 1/\hat{\sigma}_i^2$.

- Although the $\sigma_i^2$ are estimated (since they are not known), any effect is ignored in practice.
Random effects models

- DerSimonian and Laird (1986).
  - see also Brockwell and Gordon (2001).
- Extend fixed effects model to allow between-study heterogeneity.
- Study $i$, $i = 1, \ldots, k$

\[
Y_i = \theta_i + e_i \quad \text{where } e_i \sim \mathcal{N}(0, \sigma_i^2)
\]

\[
\theta_i = \mu + \epsilon_i \quad \text{where } \epsilon_i \sim \mathcal{N}(0, \tau^2)
\]

- $\tau^2$ is the between-study variation.
- Estimate of $\mu$ similar to weighted average outlined above:

\[
\hat{\mu} = \frac{\sum w_i(\tau)Y_i}{\sum w_i(\tau)} \quad \text{and} \quad \widehat{\text{var}}[\hat{\mu}] = \frac{1}{\sum w_i(\tau)}
\]

where $w_i(\tau) = 1/(\tau^2 + \hat{\sigma}_i^2)$.
- Require an estimate of $\tau^2$:
  - may be obtained by considering $Q_w$ from Cochran's $\chi^2$ test for heterogeneity
  - specifically, note that

\[
E[Q_w] = k - 1 + \tau^2 \left( \sum w_i - \frac{\sum w_i^2}{\sum w_i} \right).
\]
• solving for $\tau^2$ gives

$$\hat{\tau}^2 = \frac{Q_w - (k - 1)}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}}$$

• referred to as a method of moments estimator.
• note: $\hat{\tau}^2$ may be negative, so take $\text{max}(0, \hat{\tau}^2)$.

• Iburpofen: $\hat{\tau}^2 = 0.53$.

• Maximum likelihood-based approaches are possible:
  • summarised by Brockwell and Gordon (2001).

• Ibuprofen:
  
  Fixed effects: 7.73 95% CI (6.29, 9.51)
  Random effects: 7.96 95% CI (5.52, 11.47)

• point estimates are similar.
• confidence intervals are wider.

• Two main issues:
  • normality assumption may be problematic.
  • do not take into account uncertainty associated with $\hat{\tau}^2$. 
Bayesian models

- Variety of issues associated with fixed/random effects models.

- Asymptotic theory is the basis for validity:
  - standard error estimates and confidence intervals only valid for ‘large’ samples.
  - simulation studies of Brockwell and Gordon (2001) suggest problems.

- No account for uncertainty associated with ‘plug-in’ estimator of $\tau^2$.

- General difficulty in interpreting p-values and confidence intervals.

- Bayesian statistical paradigm offers an alternative:
  - incorporate prior information (if available).
  - account for all uncertainty (across all parameters) in the model.
  - attractive interpretations of inference.
  - flexible treatment of functions of parameters.
  - prediction, say for a new study, is easy.

- Limited use in the past due to computational issues:
  - methodological advances, e.g. MCMC.
  - more powerful computers.
Three-stage hierarchical model

- Stage 1: model for the observed data

\[ r_i^C \sim \text{Binomial}(n_i^C, p_i^C) \]
\[ r_i^T \sim \text{Binomial}(n_i^T, p_i^T) \]

- Stage 2: assumptions regarding outcome probabilities

\[ \logit(p_i^C) = \delta_i \]
\[ \logit(p_i^T) = \delta_i + \mu_i \]

  - intercepts: subject-specific \( \delta_i \)'s
  - slope: \( \mu_i \sim \mathcal{N}(\mu, \tau^2) \).

- Stage 3: Prior assumptions

\[ \delta_i \sim \mathcal{N}(0, 1000000) \]
\[ \mu \sim \mathcal{N}(0, 1000000) \]
\[ \tau^2 \sim \text{Inverse Gamma}(0.5, 0.005) \].

- Ibuprofen common odds ratio:

  Fixed effects: 7.73 95% CI (6.29, 9.51)
  Random effects: 7.96 95% CI (5.52, 11.47)
  Bayesian model: 10.52 95% CI (6.94, 16.85)

- Between-study variance: 0.53 vs 0.89 (95% CI: 0.37, 2.04).
**Additional comments**

- Majority of work to be done before employing statistical methods.

- Relatively little what can be done with statistical methodology:
  - statistical tests tend to have low power and should be interpreted with caution.
  - false impression of precision of estimation.
  - statistical heterogeneity vs clinical heterogeneity.

- Bayesian approach:
  - easily extended to other effect measures (e.g. relative difference).
  - easily extended to include additional covariates (e.g. study design indicator, dose information).
Software

- Stata
  - user-written code available through Stata Technical Bulletin.
  - type: ‘update all, search meta’.
  - Egger, Smith, and Altman (2001). *Systematic Reviews in Health Care: Meta-analysis in Context*

- SAS
  - Arthur, Bennett, and Huffcutt (2001). *Conducting Meta-Analysis using SAS.*
  - PROC MEANS

- R
  - free and open source.
  - two contributed packages; ‘meta’ and ‘rmeta’.
  - plots and analyses for this talk use R (code available).

- WinBugs
  - free and open source.
  - numerous meta-analysis examples included in the distribution.
  - code for ibuprofen example available.
Literature


