Meta Analysis

David R Urbach MD MSc
Outcomes Research Course
December 4, 2014
Overview

• Definitions
  • Identifying studies
  • Appraising studies
  • Quantitative synthesis
  • Presentation of results
  • Examining heterogeneity
  • Bias
Reviews

• Narrative Review
  – Qualitative, narrative summary by “expert”
  – Informal and subjective methods used to collect and interpret information

• Systematic review
  – Comprehensive search for relevant studies
  – Studies are appraised and synthesized according to a predetermined and explicit method

• Meta analysis
  – The statistical combination of 2 or more studies to produce a single estimate of the effect of a health intervention
Review Articles

- Meta analyses
- Systematic reviews
- Reviews
Quantitative Synthesis

• Systematic reviews do not require the statistical pooling of data from individual studies

• Sometimes it is best not to pool data
  – Quality of individual studies
  – Heterogeneity
    • Study designs
    • Interventions
    • Outcomes
    • Results
Overview

• Definitions
• Identifying studies
• Appraising studies
• Quantitative synthesis
• Presentation of results
• Examining heterogeneity
• Bias
Search Strategy

- Well-formulated clinical question
- Inclusion and exclusion criteria
- Databases
- Electronic search algorithm
Inclusion and Exclusion Criteria

• Type of study design
• Type of publication
  – Abstracts, unpublished
• Language
• Publication year
• Population
• Interventions
• Outcomes
Overview

- Definitions
- Identifying studies
- Appraising studies
- Quantitative synthesis
- Presentation of results
- Examining heterogeneity
- Bias
Why to Measure Study Quality

• Determine whether to include a study
• Determine if differences in study quality explains heterogeneity in results
• Weight a study’s results by quality
• To provide a sense of the strength of the evidence
Overview

• Definitions
• Identifying studies
• Appraising studies

• Quantitative synthesis
• Presentation of results
• Examining heterogeneity
• Bias
Effect Measures

Outcome

- Binary
  - Odds Ratio (OR)
  - Relative Risk (RR)
  - Risk Difference (RD)

- Continuous
  - Mean Difference (MD)
  - Standardized Mean Difference (SMD)
  - Weighted Mean Difference (WMD)

Overall Estimate
- Fixed Effects
- Random Effects

Overall Estimate
- Fixed Effects
- Random Effects
Dichotomous Outcomes

- Odds Ratio
- Relative Risk
- Risk Difference

Logarithmic Scale:
Dichotomous Data

$P_e = \text{event rate in experimental group}$
$P_c = \text{event rate in control group}$

- RD = risk difference $= P_e - P_c$
- RR = relative risk $= \frac{P_e}{P_c}$
- OR = odds ratio $= \frac{P_e/(1-P_e)}{[P_c/(1-P_c)]}$
Dichotomous Data

Experimental event rate = 40%
Control event rate = 30%

- RD = 40% – 30% = 10%
- RR = 40% / 30% = 1.3
- OR = (40%/60%)/(30%/70%) = 1.56
Dichotomous Outcomes

• Relative risk ~ odds ratio
  – When outcomes are uncommon
• Risks easier to interpret than odds
• Odds ratios have better mathematical properties than relative risks
• RR better when outcomes are common
• Peto OR better when outcomes are very rare
Continuous Outcomes

• Mean difference
• Standardized mean difference
# Continuous Data

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>$n_e$</td>
<td>$\bar{X}_e$</td>
<td>$s_e$</td>
</tr>
<tr>
<td>Control</td>
<td>$n_c$</td>
<td>$\bar{X}_c$</td>
<td>$s_c$</td>
</tr>
</tbody>
</table>

Mean Difference (MD) = $\bar{X}_e - \bar{X}_c$

Standardized Mean Difference (SMD) $d = \frac{\bar{X}_e - \bar{X}_c}{S}$
Example: Length of Stay

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
<th>Mean (days)</th>
<th>Standard deviation (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laparoscopic</td>
<td>50</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Open</td>
<td>50</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

Mean difference = 6 – 4 = 2 days

Standardized mean difference = (6 – 4)/2 = 1
Mean Difference vs. Standardized Mean Difference

• Mean difference
  – Comparable outcome measures
    • Scale, length of follow up
  – Meta analysis: “Weighted mean difference”

• Standardized mean difference
  – Different outcome measures for similar outcome
    • VAS, pain score
  – Converts scores to common scale
    • Number of standard deviations
Estimation of a Common Effect

• Weighting of individual studies
• Weighted by *precision* of the study
  – Dichotomous outcome
    • Number of events
  – Continuous outcome
    • Variation (standard deviation)
• Fixed and random effects
Fixed Effects Model

• One single average effect
• All studies are drawn from a population of studies measuring this effect
  – “Fixed effect”
  – Variation between studies due to chance alone
Fixed Effects
Random Effects Model

• No single “true” effect
• True effect varies according to a distribution
• Two sources of variation
  – Within studies
  – Between studies
Random Effects
Fixed vs. Random Effects

• Random effects model more “conservative”
  – Wider confidence intervals
• Gives more weight to smaller studies than fixed effects model
• Often yields similar results
  – Especially when there is little variation
• If results are markedly different, re-consider combining data!
Fixed Effects Model

• From each study
  – Effect estimate
  – Variance of effect estimate

• Combined by \textit{weighted average}:

\[
\text{Pooled Estimate} = \frac{\sum \text{estimate} \times \text{weight}}{\sum \text{weights}}
\]

Where weight = \frac{1}{\text{variance of estimate}}

• Assumes a common underlying effect estimated by every trial
Cochran’s Chi-Square Test of Heterogeneity (Q test)

\[ \chi^2_{\text{homogeneity}} = \sum_{i} W_i (Y_i - \hat{\theta})^2 \approx \chi^2_{k-1} \]

- ‘Large’ = heterogeneity
- A conservative test
  - P<0.10 indicates “statistically significant” heterogeneity
Random Effects Model

- Assumes true effect varies across studies
- Two sources of variation
  - Within studies
  - Between studies ("heterogeneity")
- Weight = 1/(variance + heterogeneity)
- When there is heterogeneity:
  - May obtain a different pooled estimate
  - Wider confidence interval
  - Larger P value
Fixed vs. Random Effects

- **Weighting** of each study:

  \[
  \text{Weight} = \frac{1}{\text{Variance}} \quad \text{Fixed Effects Model} \\
  \text{Weight} = \frac{1}{\text{Variance} + \text{Heterogeneity}} \quad \text{Random Effects Model}
  \]
Overview

- Definitions
- Identifying studies
- Appraising studies
- Quantitative synthesis
- Presentation of results
- Examining heterogeneity
- Bias
Graphical Presentation

• Forest plot
  – Each trial
    • Point estimate (square)
    • 95% confidence interval
    • Size of square indicates relative weight of study
  – Solid vertical line (“no effect”)
    • OR, RR=1.0
    • Mean difference=0
  – Diamond indicates combined estimate and 95% CI
Inguinal Hernia Repair: Recurrence

Review: Open mesh versus non-mesh for groin hernia repair
Comparison: 01 Open mesh versus open non-mesh
Outcome: 12 Recurrence

<table>
<thead>
<tr>
<th>Study</th>
<th>Mesh n/N</th>
<th>Non-mesh n/N</th>
<th>Peto Odds Ratio</th>
<th>Weight (%)</th>
<th>Peto Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 Flat mesh versus non-mesh</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barcelona 1997</td>
<td>1/140</td>
<td>36/273</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belgium 1998</td>
<td>1/93</td>
<td>2/90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bucharest 2000</td>
<td>2/130</td>
<td>4/100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copenhagen 1998</td>
<td>5/102</td>
<td>10/100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Halsted 1999</td>
<td>0/83</td>
<td>9/80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lansing 1990</td>
<td>2/71</td>
<td>7/37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lisbon (mnpb)</td>
<td>2/176</td>
<td>2/102</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Hampshire 1988</td>
<td>0/54</td>
<td>0/50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Octavia (mnpb)</td>
<td>0/149</td>
<td>7/147</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Padova 1990</td>
<td>0/198</td>
<td>0/157</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pantyrid 1988</td>
<td>0/42</td>
<td>0/38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rome 1995</td>
<td>0/53</td>
<td>1/50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rotterdam 1998</td>
<td>1/146</td>
<td>12/140</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voorhees 1984</td>
<td>1/132</td>
<td>0/107</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voorhees 2000</td>
<td>1/134</td>
<td>5/103</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subtotal
Test for heterogeneity: ch-square=6.00 df=11 p=0.96 I^2=0.0%
Test for overall effect z=0.87 p=0.39

02 Plug and mesh versus non-mesh

<table>
<thead>
<tr>
<th>Study</th>
<th>Mesh n/N</th>
<th>Non-mesh n/N</th>
<th>Peto Odds Ratio</th>
<th>Weight (%)</th>
<th>Peto Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berlin 1990</td>
<td>0/69</td>
<td>0/60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glinsk 1997</td>
<td>7/73</td>
<td>14/61</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subtotal
Test for heterogeneity: not applicable
Test for overall effect z=2.01 p=0.04

03 Perforated mesh versus non-mesh

<table>
<thead>
<tr>
<th>Study</th>
<th>Mesh n/N</th>
<th>Non-mesh n/N</th>
<th>Peto Odds Ratio</th>
<th>Weight (%)</th>
<th>Peto Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOUR 1997</td>
<td>13/100</td>
<td>4/207</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subtotal
Test for heterogeneity: not applicable
Test for overall effect z=2.31 p=0.02

Total
Test for heterogeneity ch-square=30.07 df=13 p=0.004 I^2=57.5%
Test for overall effect z=5.87 p=0.00001

Scott NW et al. Cochrane Review 2001
Overview

• Definitions
• Identifying studies
• Appraising studies
• Quantitative synthesis
• Presentation of results
• Examining heterogeneity
• Bias
Heterogeneity

Random Effects Analysis

Confidence interval of summary measure
Subgroup Analysis

Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy

133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women

Early Breast Cancer Trialists' Collaborative Group
Overview

• Definitions
• Identifying studies
• Appraising studies
• Quantitative synthesis
• Presentation of results
• Examining heterogeneity
• Bias
Bias

• Bias within studies
  – Internal validity
    • Randomization
    • Concealed allocation
    • Blinding

• Publication bias
  – Tendency for “negative” studies to remain unpublished
Summary

• Systematic review of the literature
• Estimate treatment effect in light of conflicting data
• Increase statistical power of small “underpowered” studies
• Understand variations in the results of different studies