Meta-analytical methods to identify who benefits most from treatments:

*Daft, deluded, or deft* approach?

David Fisher

ICTM Key Advances in Clinical Trials, UCL

19th May 2017
Meta-analysis

**Definition:**

“The statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings”

– Glass (1976)

- Ideally embedded within a **systematic review**
**Stratified medicine**

**Definition:**

“…identifying subgroups of patients with distinct mechanisms of disease, or particular responses to treatments… [to] ensure that the right patient gets the right treatment at the right time”

– MRC website

- Major interdisciplinary, collaborative research area
- Statistically, need to identify **subgroup interactions**
- Meta-analysis can help with this, but issues arise that do not apply to single trials in isolation
Ecological bias

• “Ecological studies” are a type of observational study
  − Units of analysis are groups, not individuals
  − Cannot make inferences about individuals based on the group to which they belong
  − Doing so is the “ecological fallacy”

• Similar issues can arise in meta-analysis, with subgroups of individuals
  For example:
  “Studies with more males than females showed a better overall treatment effect”
  does NOT necessarily imply:
  “Males benefit more from treatment than females”
Graphical illustration of “ecological bias”

- Within-trial evidence
- Across-trial evidence

Proportion male

Treatment effect on change in SBP (mm Hg)
Daft, deluded or deft?

- The source of ecological bias is use of “across-trial” (c.f. ecological group-level) evidence to inform “within-trial” (c.f. individual-level) conclusions

- Hence:
  - *Daft* = “across-trial” only
  - *Deluded* = “across-trial” and “within-trial” combined
  - *Deft* = “within-trial” only
Examples of subgroup forest plots
Ordering matters!

“By subgroup within trial”

“By trial within subgroup”

Subgroup and Trial

<table>
<thead>
<tr>
<th>Subgroup and Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carer present</td>
</tr>
<tr>
<td>Adelaide 2000</td>
</tr>
<tr>
<td>Belfast 2004</td>
</tr>
<tr>
<td>London 1999</td>
</tr>
<tr>
<td>Manchester 2001</td>
</tr>
<tr>
<td>Montreal 2000</td>
</tr>
<tr>
<td>Newcastle 1997</td>
</tr>
<tr>
<td>Oslo 2000</td>
</tr>
<tr>
<td>Stockholm 1998</td>
</tr>
<tr>
<td>Trondheim 2004</td>
</tr>
</tbody>
</table>

Subgroup (I-squared = 61.5%)

<table>
<thead>
<tr>
<th>Subgroup and Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>No carer</td>
</tr>
<tr>
<td>Adelaide 2000</td>
</tr>
<tr>
<td>Belfast 2004</td>
</tr>
<tr>
<td>London 1999</td>
</tr>
<tr>
<td>Manchester 2001</td>
</tr>
<tr>
<td>Montreal 2000</td>
</tr>
<tr>
<td>Newcastle 1997</td>
</tr>
<tr>
<td>Oslo 2000</td>
</tr>
<tr>
<td>Stockholm 1998</td>
</tr>
<tr>
<td>Trondheim 2004</td>
</tr>
</tbody>
</table>

Subgroup (I-squared = 3.6%)
“Is there really a problem?”
– Literature review

• We searched Medline (2011-14) for IPDMA with ≥1 interaction/subgroup investigation
• We extracted primary method of interaction/subgroup analysis, and of presentation
• If sufficient details available, we re-analysed data using deft, daft & deluded approaches
  • Assessed evidence for deft and daft being in agreement
  • Assessed evidence for bias in deluded estimates relative to deft
Results: Literature review

• 184 unique results; 82 eligible
• 2% deft; 23% deluded; 0% daft
• 66% unclear!
  – of which 83% used “one-stage” models
• 56% used binary subgroups
  – even when underlying data were continuous or ordinal
• 43% presented forest plot by subgroup only
  – a further 9% presented by trial within subgroup
Results: Re-analysis of published data

• 6 MAs from our review presented sufficient data for re-analysis
  – Giving 31 distinct analyses

• “Within-only” (deft) & “across-only” (daft) effects were poorly correlated
  – $\rho = 0.16$, $p = 0.39$

• Combining within & across (deluded) increased chances of a statistically significant result
  – 16% vs 6% in our sample, $p = 0.022$

• No systematic bias in effect size for deluded vs deft
  – But substantial variation (±20% in our sample)
Scatter plot of *deft* vs *daft* p-values

- **Consistent direction of effect**
- **Inconsistent direction of effect**
Results: Re-analysis of published data

- 6 MAs from our review presented sufficient data for re-analysis
  - Giving 31 distinct analyses
- “Within-only” (deft) & “across-only” (daft) effects were poorly correlated
  - $\rho = 0.16$, $p = 0.39$
- Combining within & across (deluded) increased chances of a statistically significant result
  - 16% vs 6% in our sample, $p = 0.022$
- No systematic bias in effect size for deluded vs deft
  - But substantial variation (±20% in our sample)
Bland-Altman plot: do *deft* and *deluded* estimates systematically differ?

Average effect size (Haz. Ratio or Odds Ratio) vs. % difference in effect size: 
- *p*(deluded) > 0.1 and *p*(deft) > 0.1
- *p*(deluded) < 0.1 and *p*(deft) < 0.1
- *p*(deluded) < 0.1 and *p*(deft) > 0.1
- *p*(deluded) < 0.05 and *p*(deft) > 0.05
Conclusions & Recommendations

- “Deluded” analysis does have potential real-world implications
- Poor reporting and presentation is a key shortcoming of MA subgroup analysis
  - especially combined with inappropriate methods
- “Deft” analysis and presentation should become standard practice
  - Least risk of bias
  - Clearly presented & easily interpreted via forest plots (ipdmetan)
- Trialists should routinely report all subgroup analyses pre-specified in SAP
  - in a Web Appendix if necessary; no need to worry about multiple testing
### Forest plot of interactions using **ipdmetan** in Stata

#### Trial

<table>
<thead>
<tr>
<th>Trial</th>
<th>Diff. in mean</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>diffs. (95% CI)</td>
<td>Weight</td>
</tr>
<tr>
<td>Adelaide 2000</td>
<td>6.50 (-12.65, 25.65)</td>
<td>14.05</td>
</tr>
<tr>
<td>Belfast 2004</td>
<td>-29.50 (-61.82, 2.82)</td>
<td>4.93</td>
</tr>
<tr>
<td>London 1999</td>
<td>11.90 (-4.51, 28.31)</td>
<td>19.13</td>
</tr>
<tr>
<td>Manchester 2001</td>
<td>63.90 (4.21, 123.59)</td>
<td>1.45</td>
</tr>
<tr>
<td>Montreal 2000</td>
<td>(Insufficient data)</td>
<td></td>
</tr>
<tr>
<td>Newcastle 1997</td>
<td>7.50 (-18.70, 33.70)</td>
<td>7.51</td>
</tr>
<tr>
<td>Oslo 2000</td>
<td>-4.90 (-20.11, 10.31)</td>
<td>22.26</td>
</tr>
<tr>
<td>Stockholm 1998</td>
<td>15.80 (2.15, 29.45)</td>
<td>27.66</td>
</tr>
<tr>
<td>Trondheim 2004</td>
<td>-0.90 (-42.23, 40.43)</td>
<td>3.02</td>
</tr>
<tr>
<td>Overall (I-squared = 45.3%)</td>
<td>6.47 (-0.70, 13.65)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

#### Diff. in mean

- **Favours greater effect of ESD with carer present**
- **Favours greater effect of ESD with no carer present**
Further work

• Are there any circumstances where “deluded” results might be preferred?
  - does it depend on the question being asked?
• Can we derive subgroup treatment effects consistent with the “deft” interaction?
  - what assumptions would we be making?
• How do these ideas generalise to network meta-analysis?
  - what is an “indirect” interaction effect?
  - can within & across effects be separated?
  - how to handle inconsistency?
Thank you!
d.fisher [at] ucl.ac.uk

• Colleagues:
  – Jayne Tierney, James Carpenter, Tim Morris, Ian White; MRC Clinical Trials Unit at UCL
  – Suzanne Freeman; University of Leicester

• References: Interactions