AN EMPIRICAL INVESTIGATION OF THE IMPACT OF DIFFERENT METHODS FOR SYNTHESISING EVIDENCE IN A NETWORK META-ANALYSIS

Project team
• Amalia (Emily) Karahalios - School of Public Health and Preventive Medicine, Monash University, Australia
• Simon Turner – School of Public Health and Preventive Medicine, Monash University, Australia
• Joanne McKenzie – School of Public Health and Preventive Medicine, Monash University, Australia
• Peter Herbison – University of Otago, Dunedin, New Zealand
• Georgia Salanti – Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland
• Ian White – MRC Biostatistics Unit, Cambridge, UK and MRC Clinical Trials Unit at UCL, London, UK
• Areti Angeliki Veroniki – Li Ka Shing Knowledge Institute, St. Michael’s Hospital, Toronto, Canada
• Adriani Nikolakopoulou – Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

Funding
Aim/methods

• Assess impact of re-analysing published NMAs with binary outcomes using contrast-synthesis and arm-synthesis models

• Investigate results w.r.t. characteristics of the NMA (not presented here)
  – # treatments: # studies
  – # treatments: # comparisons
  – # studies: # treatments
  – proportion of arms with <10 events/outcomes
Eligibility criteria

• We included a subset of networks from a database of networks of randomised trials (Petropolou et al 2016)

• Our subset included networks meeting the following criteria:
  – Primary outcome was binary
  – No evidence of inconsistency
  – Outcome data available

Flowchart of networks included in analysis

456 networks

272 excluded:
- No or incomplete outcome data
- No binary outcome

184 eligible networks with outcome data

26 excluded:
- 3 contained missing data
- 23 p-value of design by treatment less than 0.10

158 networks
<table>
<thead>
<tr>
<th>Method label</th>
<th>Package used in R</th>
<th>Contrast-level or arm-level input data</th>
<th>Frequentist or Bayesian framework</th>
<th>Likelihood and link functions</th>
<th>Heterogeneity</th>
<th>Treatment specific fixed effects</th>
<th>Mean effect of treatment k relative to baseline</th>
<th>Heterogeneity or random effects parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast-synthesis model 1</td>
<td>gemtc (version 0.8.1)</td>
<td>Arm-level</td>
<td>Bayesian</td>
<td>Binomial likelihood and logit link</td>
<td>Homogeneous/ common</td>
<td>N/A</td>
<td>$d_k \sim N(0, (15*5)^2)$</td>
<td>$\tau_{bk} \sim U(0,10)$</td>
</tr>
<tr>
<td>Contrast-synthesis model 2</td>
<td>gemtc (version 0.8.1)</td>
<td>Arm-level</td>
<td>Bayesian</td>
<td>Binomial likelihood and logit link</td>
<td>Homogeneous/ common</td>
<td>N/A</td>
<td>$d_k \sim N(0, (15*5)^2)$</td>
<td>Informative</td>
</tr>
<tr>
<td>Contrast-synthesis model 3</td>
<td>netmeta (version 0.9.2)</td>
<td>Contrast-level</td>
<td>Frequentist</td>
<td>N/A</td>
<td>Homogeneous/ common</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Arm-synthesis model 1</td>
<td>pcnetmeta (version 2.4)</td>
<td>Arm-level</td>
<td>Bayesian</td>
<td>Binomial likelihood and probit link</td>
<td>Homogeneous/ common</td>
<td>$\mu_k \sim N(0, 1000)$</td>
<td>N/A</td>
<td>$\sigma_k \sim U(0,10)$</td>
</tr>
<tr>
<td>Arm-synthesis model 2</td>
<td>pcnetmeta (version 2.4)</td>
<td>Arm-level</td>
<td>Bayesian</td>
<td>Binomial likelihood and probit link</td>
<td>Heterogeneous</td>
<td>$\mu_k \sim N(0, 1000)$</td>
<td>N/A</td>
<td>$\sigma_k \sim U(0,10)$</td>
</tr>
</tbody>
</table>
Preliminary results

Using graphical displays, we have compared estimates of the following parameters between the four models:

- log(OR)
- standard error(log(OR))
- ranks derived from SUCRA values
Flowchart of networks analysed

158 networks available

7 networks had 1 or more treatment arm that failed to run using arm-synthesis model 1

151 eligible networks with outcome data

31 networks failed to converge using one or more of the Bayesian methods*:
  contrast-synthesis model 1: 11
  contrast-synthesis model 2: 13
  arm-synthesis model 1: 25

120 networks available for analysis

*Numbers do not sum to 31 because some networks failed to converge for more than one model
Time taken after excluding the networks that failed to converge \((n = 120)\)

<table>
<thead>
<tr>
<th>Model</th>
<th>Average</th>
<th>SD</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast-synthesis model 1</td>
<td>5.24</td>
<td>4.32</td>
<td>4.00</td>
<td>1.00</td>
<td>20.00</td>
</tr>
<tr>
<td>Contrast-synthesis model 2</td>
<td>5.26</td>
<td>4.40</td>
<td>4.37</td>
<td>0.00</td>
<td>19.66</td>
</tr>
<tr>
<td>Contrast-synthesis model 3</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Arm-synthesis model 1</td>
<td>98.89</td>
<td>202.49</td>
<td>37.14</td>
<td>6.55</td>
<td>1262.66</td>
</tr>
</tbody>
</table>

*Note that all times measured in minutes*
Comparison of the effect estimates and standard errors
2 studies
Treatment 1: 2 events, 81 participants
Treatment 7: 61 events, 172 participants
Comparison of the ranks and SUCRA values between methods
Summary

- From our preliminary results:
  - Good agreement between the contrast-synthesis methods in terms of effect estimates and treatment ranks
  - Differences are apparent in the effect estimates and ranks when comparing the arm-synthesis model to the contrast-synthesis models
  - Contrast-synthesis models have larger standard errors compared to the arm-synthesis models
  - More variability with respect to the standard errors for the arm-synthesis models compared to the other models

- Next steps:
  - Examine another arm-synthesis model
  - Fit multilevel models to estimate the differences between the methods and to explore the factors that might explain the differences