A bivariate random effects meta-analysis using copulas

Yusuke Yamaguchi
*Data Science, Global Development, Astellas Pharma Inc.*

Richard D. Riley
*Research Institute of Primary Care and Health Sciences, Keele University*
Agenda

- Introduction: bivariate meta-analysis
- Overview of copula modeling
- Bivariate random effects meta-analysis using copulas
- Simulation study and application
- Conclusions and limitations
Introduction

- We are often interested in dealing with multiple variables in the context of meta-analysis.
  - Multiple endpoints / outcomes
  - Multiple time-points
  - Multiple treatment groups / doses
  - Multivariate meta-analysis
    (Van Houwelingen et al., 2002; Jackson et al., 2011)

- We here focus on a simple situation where there are two outcomes of interest in the meta-analysis.
Bivariate meta-analysis

- Joint synthesis of two outcomes whilst incorporating their correlation.
- The bivariate meta-analysis has noticeable advantages compared with separate univariate meta-analyses for each outcome.
  - Improving the precision of pooled estimates
    (Riley et al., 2007; Riley, 2009)
  - Reduction of selective outcome reporting biases
    (Williamson et al., 2005; Kirkham et al., 2012)
- However, in general, the bivariate meta-analysis requires bivariate normality assumption of the effect sizes.
Objectives

● To propose a novel random effects model for bivariate meta-analysis, where a copula is used for expressing the relationship between outcomes.

● To demonstrate how the dependence structure between outcomes affects the inference of their mean effects.

● To illustrate the proposed model through an application to a real example.

*Some copula models have been considered in the meta-analysis of diagnostic accuracy studies (Kuss et al., 2014; Hoyer et al., 2015).
Bivariate random effects model

- Outcome 1 and Outcome 2:

\[
\begin{pmatrix}
  y_{i1} \\
  y_{i2}
\end{pmatrix} \sim N_2\left(\begin{pmatrix}
  \theta_{i1} \\
  \theta_{i2}
\end{pmatrix}, \begin{pmatrix}
  \sigma_{i1}^2 & r_i \sigma_{i1} \sigma_{i2} \\
  r_i \sigma_{i1} \sigma_{i2} & \sigma_{i2}^2
\end{pmatrix}\right)
\]

\[
\begin{pmatrix}
  \sigma_{i1}^2 & r_i \sigma_{i1} \sigma_{i2} \\
  r_i \sigma_{i1} \sigma_{i2} & \sigma_{i2}^2
\end{pmatrix}
\]

- The within-study correlations \((r_i)\) are required but rarely available in practice.
  - Estimating within-study covariances (Wei and Higgins, 2012)
  - Using external evidence via Bayesian approach (Bujkiewicz et al., 2013)
  - Alternative model that models an overall correlation (Riley et al., 2008)
Overview of copula modeling

- Bivariate copula is defined as a bivariate cumulative distribution function with uniform margins on [0,1].
- The copula modeling allows to identify marginal distributions separately from the dependence structure, and links them together into a density function.
Commonly used copulas

- Example: margins = standard normal distribution

- Gauss copula: Linear correlation

- Clayton copula: High correlation in left tail

- Gumbel copula: High correlation in right tail
Proposal

1. Identify the marginal distributions for each outcome, and link them by using a bivariate copula (modeling an overall correlation).
2. Estimate the parameters by a Markov chain Monte Carlo method.
3. Select an optimal copula based on deviance information criteria (DIC).

[Examples of modeling continuous outcomes]

- Margins: normal distributions  $y_{i1} \sim N(\mu_1, \sigma_{i1}^2 + \varphi_{i1}^2)$, $y_{i2} \sim N(\mu_2, \sigma_{i2}^2 + \varphi_{i2}^2)$
- Dependence structure: gauss, clayton, gumbel copula
- Normal margins + Gauss copula = Alternative model by Riley et al. (2008)
Advantages of the proposed model

- No need to specify the within-study correlations
- No need to assume a linear correlation between outcomes
- No need to assume normally distributed margins for each outcome
  - Skewed or heavy-tailed distributions
  - Beta-binomial distributions for sparse binary outcomes
- Use of external information via Bayesian approach
Simulation study

- Number of studies: \( n_1 = \{5, 15, 50\}, \ n_2 = 50 \)
- True overall mean effect: \( \mu_1 = 0.25, \ \mu_2 = 0.50 \)
- Within-study variance: Simulate from \( \sigma_i^2 \sim 0.25 \times \chi_1^2 \) on \([0.009, 0.6]\)
- Between-study variance: \( \tau_1^2 = \tau_2^2 = 0.4 \)
- Within-study correlations: Simulate from \( r_i \sim \text{Uniform} [0.79, 0.99] \)
- Between-study correlation (Kendall rank correlation): \( \tau_K = 0.7 \)
- Dependence structure of random effects distributions:
  \{Gauss copula, Clayton copula, Gumbel copula\}
Mean of 3,000 results for each measure

- Post. mean of $\mu_1$
- Post. SD of $\mu_1$
- Post. mean of Between-study corr. or overall corr.
- DIC

$n_1=5, 15, 50$
Key findings

- The separate univariate meta-analyses provided poor posterior SDs of the overall mean effect for outcome 1, especially when $n_1 = 5$.
- The bivariate meta-analysis and the Riley’s alternative model provided quite similar results in this simulation settings.
- Even under the various dependence structures, posterior means of the mean effect for outcome 1 were unbiased in each approach.
- Especially when $n_1 = 5$, in comparison with the Riley’s alternative model, the proposed model provided:
  - Larger posterior means of the overall correlation
  - Smaller posterior SDs of the mean effect for outcome 1.
Example
(Riley et al., 2004)

- Meta-analysis for evaluating a prognostic importance of MYCN in neuroblastoma.
  - 81 studies in total
  - Amplified versus non-amplified MYCN
  - Effect size: hazard ratio
  - Outcome: disease-free survival (DFS)
  - Outcome: overall survival (OS)
  - 25 studies provided just DFS
  - 39 studies provided just OS
### Example: Pooled estimates and overall correlation

<table>
<thead>
<tr>
<th></th>
<th>Univariate</th>
<th>Riley</th>
<th>Gauss</th>
<th>Clayton</th>
<th>Gumbel</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hazard ratio for DFS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post. mean (SD)</td>
<td>4.40 (0.13)</td>
<td>4.38 (0.12)</td>
<td>4.38 (0.12)</td>
<td>4.53 (0.12)</td>
<td>4.40 (0.11)</td>
</tr>
<tr>
<td>95% CI</td>
<td>(3.42, 5.72)</td>
<td>(3.50, 5.54)</td>
<td>(3.50, 5.54)</td>
<td>(3.60, 5.80)</td>
<td>(3.54, 5.51)</td>
</tr>
<tr>
<td><strong>Hazard ratio for OS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post. mean (SD)</td>
<td>5.10 (0.12)</td>
<td>5.18 (0.11)</td>
<td>5.18 (0.11)</td>
<td>5.03 (0.11)</td>
<td>5.17 (0.11)</td>
</tr>
<tr>
<td>95% CI</td>
<td>(4.04, 6.52)</td>
<td>(4.16, 6.52)</td>
<td>(4.17, 6.53)</td>
<td>(4.09, 6.29)</td>
<td>(4.15, 6.49)</td>
</tr>
<tr>
<td><strong>Overall correlation (Kendall rank correlation coefficient)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post. mean (SD)</td>
<td>-</td>
<td>0.54 (0.09)</td>
<td>0.54 (0.09)</td>
<td>0.54 (0.09)</td>
<td>0.60 (0.09)</td>
</tr>
<tr>
<td>95% CI</td>
<td>-</td>
<td>(0.33, 0.69)</td>
<td>(0.33, 0.69)</td>
<td>(0.34, 0.69)</td>
<td>(0.40, 0.73)</td>
</tr>
<tr>
<td>DIC</td>
<td>267.9</td>
<td>250.7</td>
<td>250.7</td>
<td>252.8</td>
<td>249.4</td>
</tr>
</tbody>
</table>

**Univariate**: Separate univariate meta-analyses, **Riley**: Alternative model by Riley *et al.* (2008)

**Gauss**: Proposed model with gauss copula, **Clayton**: Proposed model with clayton copula

**Gumbel**: Proposed model with gumbel copula
Conclusions and limitations

● We proposed a bivariate meta-analysis using copulas:
  ✓ Flexible modeling beyond the simplistic assumptions like the normality and the linear dependence

● Limitations of the proposed model:
  ✓ Loss of a fully hierarchical structure
  ✓ Difficulty in calculating a joint prediction region
  ✓ Modeling negatively correlated outcomes
  ✓ Selection of an optimal copula
Reference

- Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. *Statistics in Medicine* 2012; 31:2179-2195.
Reference

Thank you for your attention

yusuke-yamaguchi@astellas.com