Do Randomized and Non-Randomized Trials Yield Different Answers in Similar Populations? – Evidence from a 'Meta-Propensity Score' Analysis in Cardiac Surgery

Kuss O¹, Legler T¹, Börgermann J²

¹Institute of Medical Epidemiology, Biostatistics, and Informatics, University of Halle-Wittenberg, Halle (Saale)
²Department of Cardiac and Thoracic Surgery, Friedrich Schiller University Jena
Introduction I: RCTs and Non-RCTs

- Effects of therapeutic interventions should be checked (if possible) in randomized controlled trials (RCTs).
- RCTs sometimes have limited *external* validity (Rothwell, 2005).
- Consequence for all systematic comparisons of RCTs and Non-RCTs: Limited *internal* validity!

If RCTs are conducted in highly selected populations, but Non-RCTs in general populations, potential differences between RCTs and Non-RCTs are not necessarily due to missing randomisation. They might also arise from the different populations involved!
Introduction II: 'Meta-Randomization'?

• Ideally we would like to have a 'Meta-randomized' trial: Investigators willing to conduct a study on a specific clinical question would be randomly selected to perform a RCT or a Non-RCT:
  → Balancing of 'Meta-confounders' (All properties of investigator’s setting and patients)
  → Causal effect of randomization could be validly measured

• Technically feasible? Ethically acceptable?
Introduction III: 'Meta-Propensity Score'!

- Our Solution: Matched 'Meta-Propensity Score-Analyse'

1. Match RCTs and Non-RCTs for relevant 'Meta-confounders' (summarized by a 'Meta-Propensity Score')
2. Compare treatment effects in the 'Meta-matched' population
Introduction IV: Clinical topic

• Comparison of on- and off-pump (beating heart, no use of the heart-lung-machine) technique in coronary artery bypass grafting

• “… one of the most hotly debated and polarizing issues in cardiac surgery …” (Sellke et al., 2005).

• Public health relevance: In Germany, 51,273 (isolated) bypass surgeries were conducted in 2006, of which 10.1% were off-pump (Gummert et al., 2007).
Methods I: Studies

- Systematic search for all RCTs (all studies from the three largest, most recent systematic reviews on the topic + own MEDLINE search)
- Systematic search for all PS-analyses (Kuss et al., 2008(?))
- Inclusion criteria for studies:
  - Information given on study population and setting ('Meta-confounders')
  - Information given on at least one of 10 binary clinical in-hospital outcomes (Postoperative death, stroke, myocardial infarction, renal failure, …)
Methods II: Studies

• Structured data extraction (pretested data extraction form, two blinded reviewers (OK, TL), differences resolved by consensus with a third reviewer (JB))

• Extracted data:
  – General information (time of study, number of centers, number of patients, country, …)
  – Study population (baseline risk factors)
  – Outcomes (Absolute numbers or effect estimates for each available clinical outcome)
Methoden III: Meta-PS-Analysis

• Inclusion criterion for 'Meta-confounders': Information in at least 2/3 of all RCTs and PS-Analyses
• Simplifying assumption: Mean = Median
• If necessary: transformation of categorical 'Meta-confounders' in continuous ones under the assumption of uniform distribution in the categories
• Multiple imputation of missing values in the 'Meta-PS-Model' (SAS® PROC MI)
• 'Meta-PS model' as logistic model with continuous 'Meta-confounders' up to third order (optimal c-statistic=89.6%)
• 'Meta-matching' with an optimal matching algorithm with variable number of controls (Soledad Cepeda et al., 2006)
Methods IV: Meta-PS-Analysis

• “Reconstruction” of four-fold-tables in the PS analyses by the Di Pietrantonj method (Di Pietrantonj, 2006).
• Estimated treatment effects from RCTs and PS analyses were compared in the 'Meta-matched' sample as differences in odds ratios (with 95%-confidence intervals) with a 3-level (patients are correlated within studies, studies are correlated within matching stratum) random effects logistic regression model
• Parameter estimation by PQL (SAS® PROC GLIMMIX)
• Confidence intervals by the multivariate delta method
Results I: Studies

• Initially retrieved: 28 PS-Analyses and 51 RCTs
• 7 'Meta-confounders' had data in at least 2/3 of all RCTs and PS-Analyses
• After 'Meta-matching':
  10 PS-Analyses (25,552 patients) and
  29 RCTs (2,723 patients)
• 186 effect estimates from all clinical outcomes:
  Post-op. death (38), stroke (28), MI (27), atrial fibrillation (16), …
## Results II: Studies before 'Meta-matching'

<table>
<thead>
<tr>
<th>Meta-confounder</th>
<th>PS analyses (N=28)</th>
<th>RCTs (N=51)</th>
<th>p-value</th>
<th>Standard. diff. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td>17 (61%)</td>
<td>36 (71%)</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>Northern America</td>
<td>10 (36%)</td>
<td>5 (10%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>1 (3%)</td>
<td>10 (19%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number of centers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>18 (65%)</td>
<td>47 (92%)</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>&gt;1</td>
<td>9 (32%)</td>
<td>3 (6%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>1 (3%)</td>
<td>1 (2%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean age (years)</strong></td>
<td>65.8</td>
<td>63.1</td>
<td>0.002</td>
<td>75.1</td>
</tr>
<tr>
<td><strong>Mean prop. males (%)</strong></td>
<td>72.1</td>
<td>77.1</td>
<td>0.138</td>
<td>-37.0</td>
</tr>
<tr>
<td><strong>Mean prop. pre-op. MI (%)</strong></td>
<td>44.5</td>
<td>41.6</td>
<td>0.480</td>
<td>21.0</td>
</tr>
<tr>
<td><strong>Mean LVEF (%)</strong></td>
<td>58.8</td>
<td>62.7</td>
<td>0.033</td>
<td>-55.9</td>
</tr>
<tr>
<td><strong>Mean prop. diabetic pers. (%)</strong></td>
<td>26.2</td>
<td>24.4</td>
<td>0.595</td>
<td>13.9</td>
</tr>
</tbody>
</table>
## Results III: Studies after 'Meta-matching'

<table>
<thead>
<tr>
<th>Meta-confounder</th>
<th>PS analyses (N=10)</th>
<th>RCTs (N=29)</th>
<th>p-value</th>
<th>Standard. diff. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td>8 (80%)</td>
<td>23 (80%)</td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td>Northern America</td>
<td>1 (10%)</td>
<td>3 (10%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>1 (10%)</td>
<td>3 (10%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number of centers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>8 (80%)</td>
<td>25 (86%)</td>
<td>0.631</td>
<td></td>
</tr>
<tr>
<td>&gt;1</td>
<td>2 (20%)</td>
<td>3 (10%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>0 (0%)</td>
<td>1 (4%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean age (years)</strong></td>
<td>64.1</td>
<td>63.9</td>
<td>0.916</td>
<td>3.9</td>
</tr>
<tr>
<td><strong>Mean prop. males (%)</strong></td>
<td>80.5</td>
<td>76.9</td>
<td>0.431</td>
<td>30.5</td>
</tr>
<tr>
<td><strong>Mean prop. pre-op. MI (%)</strong></td>
<td>44.0</td>
<td>39.9</td>
<td>0.530</td>
<td>27.6</td>
</tr>
<tr>
<td><strong>Mean LVEF (%)</strong></td>
<td>61.1</td>
<td>60.7</td>
<td>0.861</td>
<td>6.8</td>
</tr>
<tr>
<td><strong>Mean prop. diabetic pers. (%)</strong></td>
<td>24.8</td>
<td>25.2</td>
<td>0.925</td>
<td>-3.7</td>
</tr>
</tbody>
</table>
Results IV: Differences of ORs (PS-RCT) in the 'Meta-matched' sample

Differences of ORs (PS-RCT) (with 95%-CI)
Conclusion I

• In our example, treatment effects from RCTs and PS analyses were very similar in a ‘Meta-matched' population, indicating a small effect of randomisation itself. (difference in ORs [95%-CI]: -0.027 [-0.119, 0.066])

• Besides 'Meta-matching':
  – Identical design of Non-RCTs (PS)
  – Identical intervention and control group,
  – Identical responses in RCTs and Non-RCTs
  – Identical length of follow-up
  – Valid outcomes used
  – Overlap (though not perfect) in observation intervals
Conclusion II

• Limitations:
  – Publication bias?
  – Simplifying assumptions too simple?
  – 'Meta-Residual Confounding'?
    (We did not conduct a 'Meta-randomized' trial!)
  – Balancing of 'Meta-Confounders' in the 'Meta-matched' sample does not assure balancing for the individual outcome!
Conclusion III

• **In the future:**
  Our study needs independent replication in a different (preferably non-surgical) setting.

Even if replicated we do not think that RCTs would be obsolete, but the current practice of excluding well conducted Non-RCTs from systematic reviews of treatment effects could at least be questioned.
References

• Rothwell PM. External validity of randomised controlled trials: 'to whom do the results of this trial apply?' Lancet 2005; 365(9453):82-93.


