

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-Analysis'
 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 and PS-analyses
- 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, ...)
- Structured data extraction (pretested data extraction form, two blinded reviewers (OK, TL), differences resolved by consensus with a third reviewer (JB))

Methods II: 'Meta-PS-Analysis'

- Inclusion criterion for 'Meta-confounders':
Information in at least 2/3 of all RCTs and PS-Analyses
- 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 III: 'Meta-PS-Analysis'

All specifications and assumptions regarding building of the 'Meta-PS-model' were determined **a priori** and **independent of outcome data!**

Methods IV: Statistical 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), ...

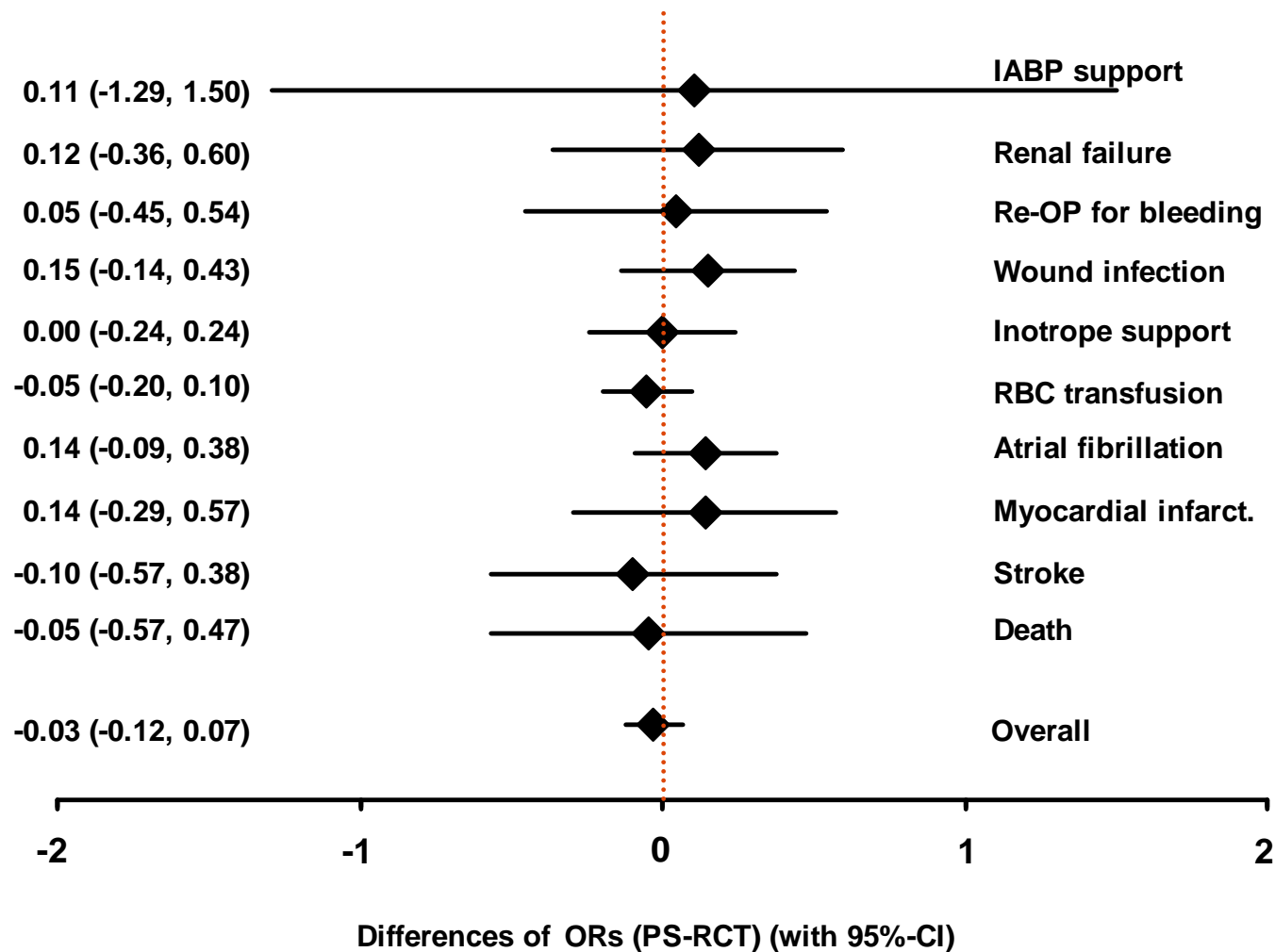
Results II: Studies *before* 'Meta-matching'

| Meta-confounder | PS analyses (N=28) | RCTs (N=51) | p-value | Standard. diff. (%) |
|-------------------------------|-----------------------|----------------|---------|------------------------|
| Study region | | | 0.007 | |
| Europe | 17 (61%) | 36 (71%) | | |
| Northern America | 10 (36%) | 5 (10%) | | |
| Others | 1 (3%) | 10 (19%) | | |
| Number of centers | | | | |
| 1 | 18 (65%) | 47 (92%) | 0.006 | |
| >1 | 9 (32%) | 3 (6%) | | |
| Missing | 1 (3%) | 1 (2%) | | |
| Mean age (years) | 65.8 | 63.1 | 0.002 | 75.1 |
| Mean prop. males (%) | 72.1 | 77.1 | 0.138 | -37.0 |
| Mean prop. pre-op. MI (%) | 44.5 | 41.6 | 0.480 | 21.0 |
| Mean LVEF (%) | 58.8 | 62.7 | 0.033 | -55.9 |
| Mean prop. diabetic pers. (%) | 26.2 | 24.4 | 0.595 | 13.9 |

Results III: Studies *after* 'Meta-matching'

| Meta-confounder | PS analyses (N=10) | RCTs (N=29) | p-value | Standard. diff. (%) |
|-------------------------------|-----------------------|----------------|---------|------------------------|
| Study region | | | 0.999 | |
| Europe | 8 (80%) | 23 (80%) | | |
| Northern America | 1 (10%) | 3 (10%) | | |
| Others | 1 (10%) | 3 (10%) | | |
| Number of centers | | | 0.631 | |
| 1 | 8 (80%) | 25 (86%) | | |
| >1 | 2 (20%) | 3 (10%) | | |
| Missing | 0 (0%) | 1 (4%) | | |
| Mean age (years) | 64.1 | 63.9 | 0.916 | 3.9 |
| Mean prop. males (%) | 80.5 | 76.9 | 0.431 | 30.5 |
| Mean prop. pre-op. MI (%) | 44.0 | 39.9 | 0.530 | 27.6 |
| Mean LVEF (%) | 61.1 | 60.7 | 0.861 | 6.8 |
| Mean prop. diabetic pers. (%) | 24.8 | 25.2 | 0.925 | -3.7 |

Results IV: Differences of ORs (PS-RCT) in the 'Meta-matched' sample



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])
- **Advantages:** (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 maybe 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 information from 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.
- Sellke FW, DiMaio JM, Caplan LR et al. Comparing on-pump and off-pump coronary artery bypass grafting: numerous studies but few conclusions: a scientific statement from the American Heart Association council on cardiovascular surgery and anesthesia in collaboration with the interdisciplinary working group on quality of care and outcomes research. *Circulation*. 2005;111:2858-2864.
- Gummert JF, Funkat A, Beckmann A, Schiller W, Hekmat K, Ernst M, Krian A. Cardiac surgery in Germany during 2006: a report on behalf of the German Society for Thoracic and Cardiovascular Surgery. *Thorac Cardiovasc Surg*. 2007;55(6):343-50.
- Soledad Cepeda M, Boston R, Farrar JT, Strom BL. Optimal matching with a variable number of controls vs. a fixed number of controls for a cohort study: trade-offs. *J Clin Epidemiol*. 2003;56(3):230-7.
- Di Pietrantonj C. Four-fold table cell frequencies imputation in meta analysis. *Statist Med* 2006; 25:2299–2322.