

Gibt es einen Unterschied zwischen randomisierten und nicht-randomisierten Studien?

**Evidenz aus einer „Meta-Propensity Score-
Analyse“ in der Herzchirurgie
(work in progress)**

Kuss O¹, Legler T¹, Börgermann J²

**¹Institut für Medizinische Epidemiologie, Biometrie und Informatik,
Medizinische Fakultät, Martin-Luther Universität Halle-Wittenberg,
Halle (Saale)**

²Klinik für Herz- und Thoraxchirurgie, Universitätsklinikum Jena

Einleitung I: RCTs und Non-RCTs

- Wirksamkeit von therapeutischen Interventionen sollte nur im Rahmen von randomisierten klinischen Studien (RCTs) geprüft werden.
- Allerdings: RCTs haben manchmal eine geringe externe Validität.
- Es gibt Situationen, wo RCTs „unnecessary, inappropriate, impossible, or inadequate“ sind (Black, 1996)
- Alternative (?): Nicht-randomisierte Studien (Non-RCTs)
- Klar: Nicht-randomisierte Therapiezuweisung kann zu verzerrten Effektschätzern führen.
- Modern: Propensity Score-Analysen (v.a. im kardiologischen Bereich, Stürmer et al., 2006)

Einleitung II: Propensity Score

- Hier: Beschränkung auf gematchte PS-Analysen
- Mehrstufiger Prozess:
 1. Stufe: Schätze Behandlungswahrscheinlichkeit für jeden Patienten
 2. Stufe: Matche Probanden mit gleicher Behandlungswahrscheinlichkeit aus Behandlungs- und Kontrollgruppe
 3. Stufe: Schätze Behandlungseffekt in gematchter Stichprobe
- Beachte Ähnlichkeit zur RCT (explizites Herstellen einer strukturgleichen Kontrollgruppe, Angabe von Behandlungsw't möglich)

Einleitung III: Unterschiede RCTs und Non-RCTs?

- Sind die Ergebnisse von RCTs und Non-RCTs zur selben Fragestellung unterschiedlich?
- Fülle von Einzelstudien, bereits gesammelt in mindestens 5 großen systematischen Reviews (Kunz et al., 2007; MacLehose et al, 2000; Britton et al. 1998; ...)
- **Fazit aus diesen:**
Ergebnisse aus nicht-randomisierten Studien unterscheiden sich manchmal, aber nicht immer, und nicht in einer vorhersagbaren Richtung von randomisierten Studien.

Einleitung IV: Unterschiede RCTs und Non-RCTs?

- **Aber:** Gehen die Unterschiede tatsächlich auf die fehlende Randomisierung zurück oder eher darauf, dass RCTs und Non-RCTs von verschiedenen Menschen in verschiedenen Populationen gemacht worden sind?

Einleitung V: Meta-Randomisation?

- Ideal wäre „Meta-Randomisierte“ Studie:
Studiengruppen, die eine Studie zu einer gewissen klinischen Frage durchzuführen bereit sind, werden zufällig ausgewählt, eine randomisierte oder eine nicht-randomisierte Studie durchzuführen:
 - Strukturgleichheit bzgl. der „Meta-Confounder“
 - Effekt der Randomisierung valide messbar
- Technisch machbar? Ethisch einwandfrei?
- Unsere Lösung:
„Meta-Propensity Score-Analyse“

Einleitung VI: Klinisches Beispiel

- Vergleich der Off-Pump-Technik (ohne Herz-Lungen-Maschine) mit der On-Pump-Technik (Standard) bei Bypass-Operationen
- “ ... one of the most hotly debated and polarizing issues in cardiac surgery...” (Sellke et al. (AHA), 2005).
- Public health-Relevanz: In Deutschland werden jährlich über 50.000 Bypass-OPs durchgeführt.

Methoden I: Studien

- Systematische Suche aller PS-Analysen (siehe Kuss et al., 2008(?))
- Systematische Suche aller RCTs (alle Studien aus 5 größten systematischen Reviews + eigene Suche)
- Einschlusskriterien für Studien:
 - Information über Studienpopulation und Studiensetting (“Meta-Confounder”)
 - Information über mindestens einen von 10 binären klinischen In-Hospital Outcomes (Tod, Schlaganfall, MI, Nierenversagen, ...)

Methoden II: Studien

- Strukturierte Datenerfassung (MS ACCESS), zwei verblindete Reviewer (OK, TL), automatischer Abgleich, Konsensusfindung mit einem dritten Reviewer (JB)
- Erhobene Daten:
 - Studiensetting (Zeitraum, Anzahl Zentren, Anzahl Probanden, Land etc.)
 - Studienpopulation (präoperative Risiko-Faktoren)
 - Therapieeffekte (LogOdds Ratios für jeden klinischen Endpunkt)

Methoden III: Meta-PS-Analyse

- Einschlusskriterien für “Meta-Confounder”:
Information in mindestens 2/3 aller RCTs und PS-Analysen
- Vereinfachende Annahmen: Mittelwert = Median
- Falls nötig: Umrechnen von kategorialen Meta-Confoundern in stetige unter Annahme von Gleichverteilung in den Kategorien
- Multiple Imputation für Fehlstellen-Korrektur im “Meta-PS-Modell”
- “Meta-PS-Modell” als generalisiertes additives Modell mit Splines für die stetigen Meta-Confounder
- Optimal matching Algorithmus mit variabler Paargröße

Methoden IV: Meta-Meta-Regression

- Messung des Effekts des Studientyps (PS vs. RCT) durch “Meta-Meta-Regression” im gematchten Sample, getrennt für die klinischen Zielgrößen

Ergebnisse I: Studien

- Initial 28 PS-Analysen und 51 RCTs eingeschlossen
- 7 “Meta-Confounder” mit Information in mehr als 2/3 der Studien für Einschluss in “Meta-PS-Modell”
- Nach “Meta-Matching”:
10 PS-Analysen (mit 25.552 Patienten) und
29 RCTs (mit 2.723 Patienten) eingeschlossen

Ergebnisse II: Studien vor „Meta-PS-Matching“

Meta-confounder	PS analys. (N=28)	RCTs (N=51)	p-value
Studienregion			0.007
Europa	17 (61%)	36 (71%)	
Nordamerika	10 (36%)	5 (10%)	
Sonstige	1 (3%)	10 (19%)	
Anzahl Zentren			0.006
1	18 (65%)	47 (92%)	
>1	9 (32%)	3 (6%)	
Fehlend	1 (3%)	1 (2%)	
Mittl. Alter (Jahre)	65.8	63.1	0.002
Mittl. Anteil Männer (%)	72.1	77.1	0.138
Mittl. Anteil Prä-op. Herzinfarkt (%)	44.5	41.6	0.480
Mittl. LVEF (%)	58.8	62.7	0.033
Mittl. Anteil Diabetiker (%)	26.2	24.4	0.595

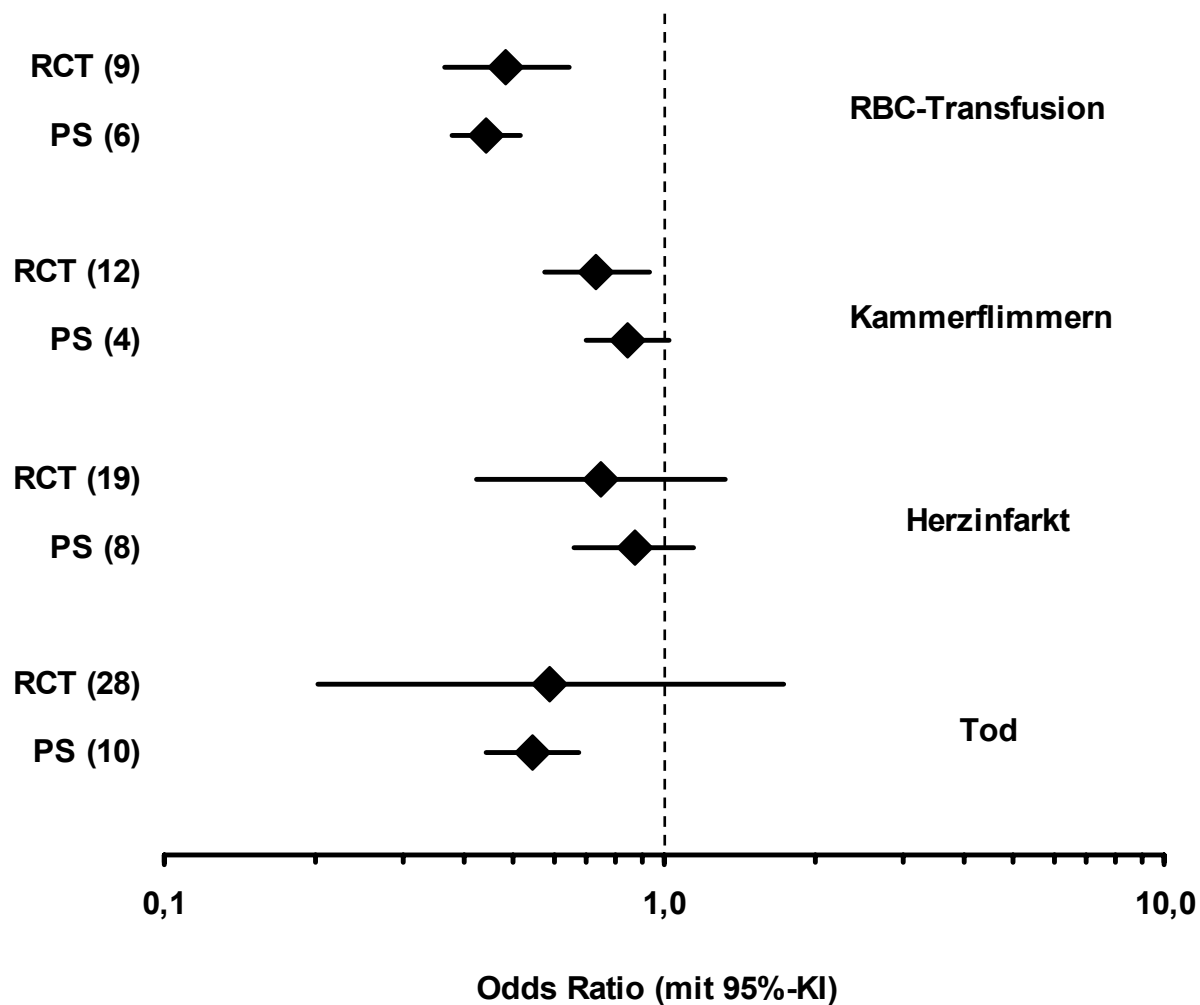
Ergebnisse II: Studien *nach* „Meta-PS-Matching“

Meta-confounder	PS analys. (N=10)	RCTs (N=29)	p-value
Studienregion			0.999
Europa	8 (80%)	23 (80%)	
Nordamerika	1 (10%)	3 (10%)	
Sonstige	1 (10%)	3 (10%)	
Anzahl Zentren			
1	8 (80%)	25 (86%)	0.631
>1	2 (20%)	3 (10%)	
Fehlend	0 (0%)	1 (4%)	
Mittl. Alter (Jahre)	64.1	63.9	0.916
Mittl. Anteil Männer (%)	80.5	76.9	0.431
Mittl. Anteil Prä-op. Herzinfarkt (%)	44.0	39.9	0.530
Mittl. LVEF (%)	61.1	60.7	0.861
Mittl. Anteil Diabetiker (%)	24.8	25.2	0.925

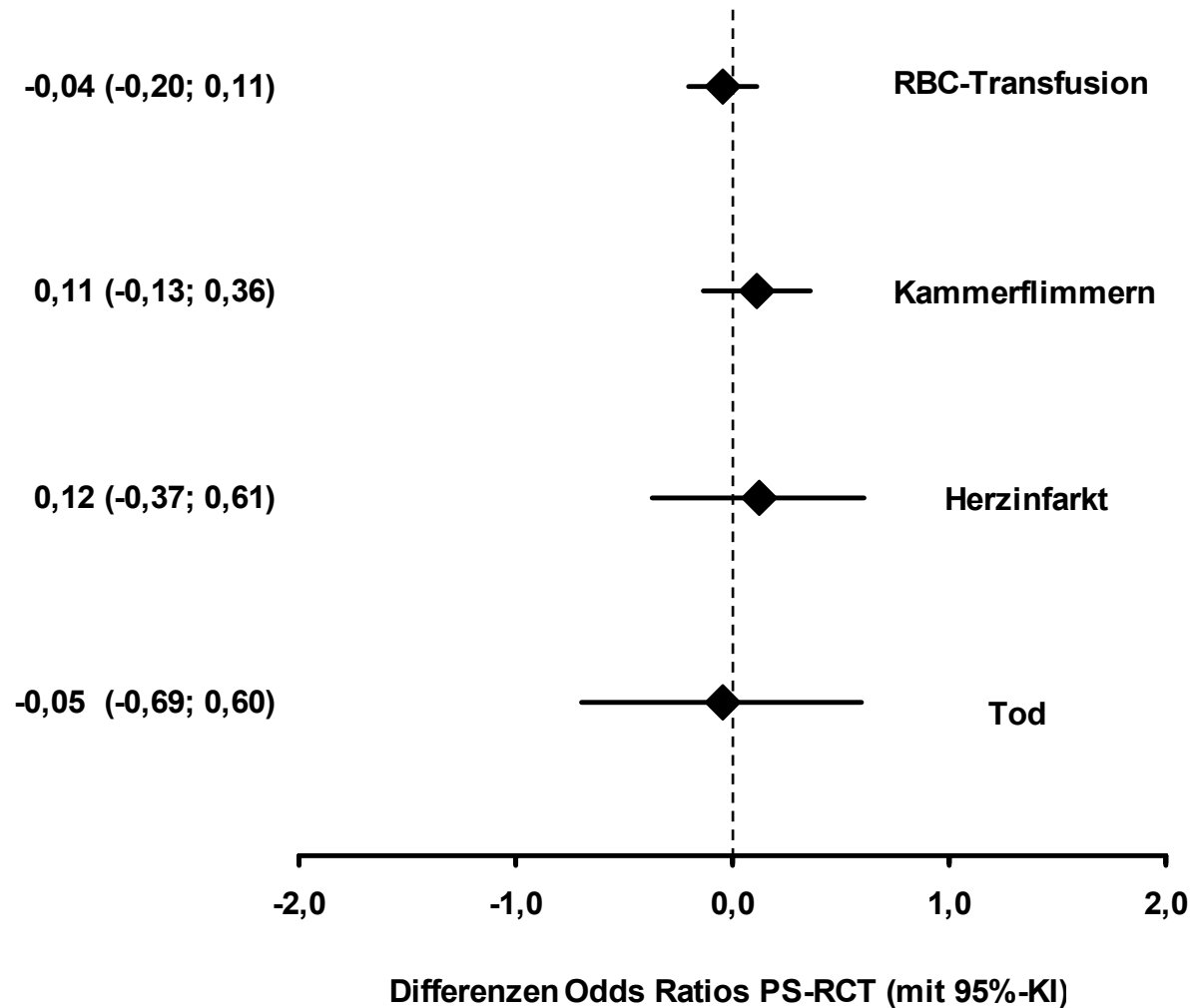
Ergebnisse IV: Klinische Zielgrößen

- 186 Effektschätzer zu den 10 klinischen Zielgrößen
- Hier Beschränkung auf 4 klinische Zielgrößen (in Klammern Anzahl Effektschätzer):
Tod (38), Herzinfarkt (27),
Kammerflimmern (16), RBC-Transfusion (11)

Ergebnisse V: MA-Schätzer nach Studientypen getrennt im gematchten Sample



Ergebnisse VI: Differenzen Odds Ratios im gematchten Sample



Fazit I

- **In unserem Beispiel:**
Sehr wenig Unterschiede zwischen RCTs und PS-Analysen bei “gleichen” Studienpopulationen
→ **Effekt der Randomisierung überschätzt?**
- **Stärken (vgl. Deeks et al.):**
 - Identisches Design der Non-RCTs (PS)
 - Identische Intervention/Kontrolle, Identische Dosierung
 - Identische Zielgrößen
 - Identische Länge des Follow-Up
 - Auswahl valider klinischer Zielgrößen
 - Überschneidung (wenn auch keine perfekte) in den zeitlichen Intervallen, in denen RCTs und PS-Analysen durchgeführt wurden.

Fazit II

- **Limitation:**
 - Publication Bias?
 - Unterschiedliche Definition der Meta-Confounder
 - „Meta-Residual Confounding“ (durch z.B. systematisch unterschiedliche Patientenkollektive)?
- **Ausblick:**
 - Re-Analyse/Rekonstruktion der individuellen Events (mit der Hoffnung auf kleinere KIs)
 - „Eine Studie ist keine Studie“

Literatur

- Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ*. 1996;312:1215-1218.
- Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol*. 2006 May;59(5):437-47.
- Britton A, McKee M, Black N et al. Choosing between randomised and non-randomised studies: a systematic review. *Health Technol Assess*. 1998;2:i-124.
- Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev*. 2007 Apr 18;(2):MR000012.
- MacLehose RR, Reeves BC, Harvey IM et al. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. *Health Technol Assess*. 2000;4:1-154.
- Deeks JJ, Dinnes J, D'Amico R et al. Evaluating non-randomised intervention studies. *Health Technol Assess*. 2003;7:iii-173.
- 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.

Ergebnisse VIII: Vergleich Gematchte (M) vs. Nichtgematchte (U) Studien

