

Gibt es einen Unterschied zwischen randomisierten und nicht-randomisierten Studien in ähnlichen Patientengruppen? - Evidenz aus einer „Meta-Propensity Score-Analyse“ in der Herzchirurgie

Kuß O¹, Legler T¹, Börgermann J²

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

²Klinik für Herz- und Thoraxchirurgie, Universitätsklinikum Jena

Einleitung I: RCTs und Non-RCTs

- Effekte von therapeutischen Interventions sollten, wenn möglich, in randomisierten kontrollierten Studien (RCTs) geprüft werden.
- RCTs haben manchmal eine geringe **externe** Validität (Rothwell, 2005).
- **Konsequenz:** Mangelnde **interne** Validität aller systematischen Vergleiche von randomisierten und nicht-randomisierten Studien!

Einleitung II: RCTs und Non-RCTs

Konkret:

Wenn **RCTs** in **hoch-selektionierten** Populationen durchgeführt werden,
Non-RCTs dagegen in **unselektionierten** Populationen,
dann sind Unterschiede nicht notwendigerweise auf die **fehlende Randomisierung** zurückzuführen.

Sie könnten auch durch die **unterschiedlichen Patientenpopulationen** zustande kommen!

Einleitung III. RCTs und Non-RCTs

- Bisherige Evidenz zum Vergleich von RCTs und Non-RCTs ist reichhaltig vorhanden und bereits in systematischen Reviews gesammelt (McKee et al., 1998; MacLehose et al., 2000; Ioannidis et al., 2001; Deeks et al., 2003; Kunz et al., 2007))
- Bisherige Evidenz ist aber auch nicht einheitlich: „ ...results from non-RCT studies differ sometimes, but not always, and not in a predictable direction from the results of RCTs”.
- Allerdings, Vermutung von Britton et al.:
“ ... that it may be possible to minimise any differences by ensuring that subjects included in each type of study are comparable.”

Einleitung IV: RCTs und Non-RCTs

Mögliche Studiendesigns

Literature review
Between-study comparison



Single study analysis
Within-study comparison

Einleitung V: RCTs und Non-RCTs

Mögliche Studiendesigns

In der
Medizin

Literature review
Between-study comparison

~~Single study analysis
Within-study comparison~~

Einleitung VI: „Meta-Randomisierung“?

- Idealerweise sollte eine „meta-randomisierte“ Studie durchgeführt werden:
Studiengruppen, die eine Studie zu einer bestimmten klinischen Frage durchzuführen bereit sind, werden zufällig ausgewählt („meta-randomisiert“), eine randomisierte oder eine nicht-randomisierte Studie durchzuführen.
 - Strukturgleichheit bzgl. aller „Meta-Confounder“
 - Kausaler Effekt der Randomisierung ist intern valide messbar.
- Technisch möglich? Ethisch akzeptabel?

Einleitung VII : „Meta-Propensity Score“!

- Unsere Lösung: Gematchte „Meta-Propensity Score-Analyse“
 1. Matche RCTs und Non-RCTs für wichtige „Meta-confounder“ (zusammengefasst durch einen „Meta-Propensity Score“)
 2. Vergleiche Effektschätzer in der “meta-gematchten” Stichprobe

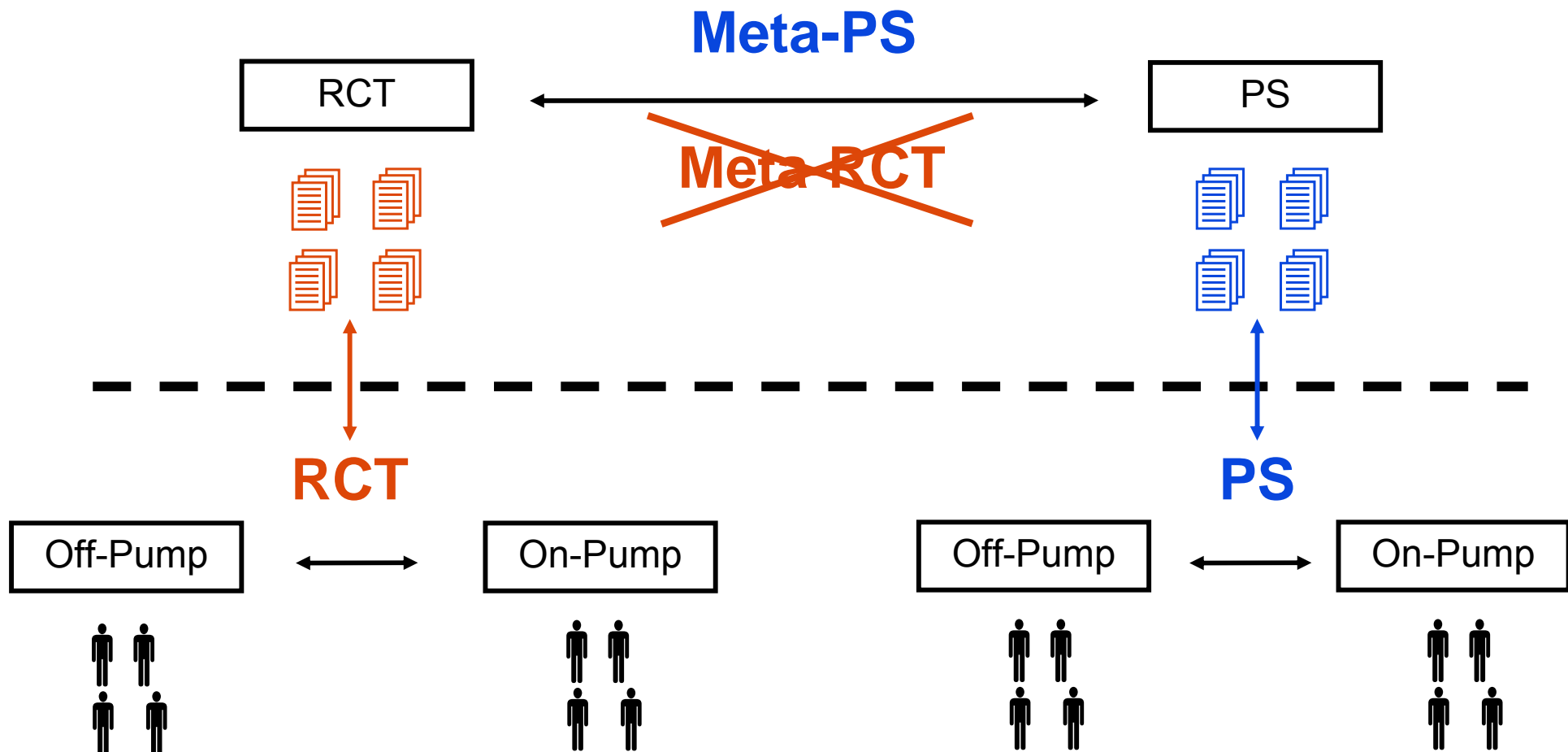
Einschränkung: In der Gruppe der Non-RCTs werden nur Propensity-Score Analysen betrachtet.

Einleitung VIII : Klinische Frage

- Vergleich zweier OP-Techniken (mit und ohne Einsatz der Herz-Lungen-Maschine, On- und Off-pump) in der Bypasschirurgie
- “ ... one of the most hotly debated and polarizing issues in cardiac surgery ...” (Sellke et al., 2005).
- Public health-Relevanz: In Deutschland wurden im Jahr 2006 51.273 (isolierte) Bypass-OPs durchgeführt, 10,1% davon „Off-pump“ (Gummert et al., 2007).

Einleitung IX: „Meta-Propensity Score“ – Versuch einer Erklärung

„Meta“-Ebene



Inhaltlich-Klinische Ebene

Material und Methoden I: Studien

- Systematische Suche nach allen RCTs (alle RCTs aus den 3 größten und aktuellsten MAs von RCT + eigene MEDLINE-Suche)
- Systematische Suche nach allen PS-Analysen (Kuss et al., 2009(?))
- Einschlusskriterien für Studien:
 - Information zu Studien-Population und anderen “Meta-Confoundern”
 - Information zu mindestens einer von 10 dichotomen klinischen In-Hospital Zielgrößen (Postoperativer Tod, Schlaganfall, Herzinfarkt, Nierenversagen, ...)

Material und Methoden II: Studien

- Strukturierte Datenerfassung (pilot-gestestetes CRF, zwei verblindete Reviewer [OK, TL], evtl. Konsensfindung mit drittem Reviewer (JB))
- Extrahierte Daten:
 - Allgemeine Angaben zur Studie (Studiendauer, Anzahl Zentren, Anzahl Patienten, Land, ...)
 - Studienpopulation (Risikofaktoren)
 - Zielgrößen (Abs. Häufigkeiten oder Effektschätzer)

Material und Methoden II: „Meta-PS-Analyse“

- Einschlusskriterium für „Meta-Confounder“:
Information in mindestens 2/3 aller RCTs und PS-Analysen
- Vereinfachende Annahme: Mittelwert = Median
- Falls notwendig: Transformation von kategorialen „Meta-Confoundern“ in stetige unter Annahme von Gleichverteilung in den Kategorien
- Multiple Imputation von fehlenden Werten im „Meta-PS-Modell“ (SAS[®] PROC MI, 1000 Datensätze)
- „Meta-PS-Modell“ als logistisches Modell mit stetigen „Meta-Confoundern“ und deren polyn. Transformationen bis zur dritten Ordnung, keine Interaktionsterme, Optimalitätskriterium: c-Statistik

Material und Methoden IV: „Meta-PS-Analyse“

- „Meta-Matching“ mit medianem (über die 1000 imputierten Datensätze) „Meta-PS“ und einem optimalen Matching-Algorithmus mit einer variablen Anzahl (höchstens 1:4) von Kontrollen (Soledad Cepeda et al., 2006)
- **Wichtig:** Festlegung des „Meta-PS-Modells“ **vor** **Auswertung** und **unabhängig vom Studien-Design**

Material und Methoden V: Auswertung

- Unterschiede zwischen RCTs und PS-Analysen werden im „meta-gematchten“ Datensatz als Differenzen von „Meta-Odds ratios“ (mit 95%-KI, multivariate Delta-Methode) angegeben.
- „Rekonstruktion“ von Vierfeldertafeln in den PS-Analysen mit der Di Pietrantonj-Method (2006, Umrechnung von Effektschätzern in zugrundeliegende Vierfeldertafeln).
- **Statistisches Modell:** Logistische Regression mit zufälligen Effekten
Für einzelne Zielgrößen: Zwei zufällige Effekte (Studie, Matching-Stratum)
Für Gesamteffekt: Zusätzlicher zufälliger Effekt Zielgröße
- Parameterschätzung durch PQL (SAS[®] PROC GLIMMIX)

Ergebnisse I: Studien

- Gefundene und eingeschlossene Studien: 28 PS-Analysen und 51 RCTs
- 7 „Meta-Confounder“ mit Information in mindestens 2/3 aller RCTs und PS-Analysen
- Nach „Meta-Matching“: 39 Studien, 10 PS-Analysen (25.552 Patienten) und 29 RCTs (2.723 Patienten)
- Insgesamt 186 Effektschätzer aus den 39 Studien für die klinischen Zielgrößen:
Postoperativer Tod: 38, Schlaganfall: 28, Herzinfarkt: 27, Nierenversagen: 12, ...

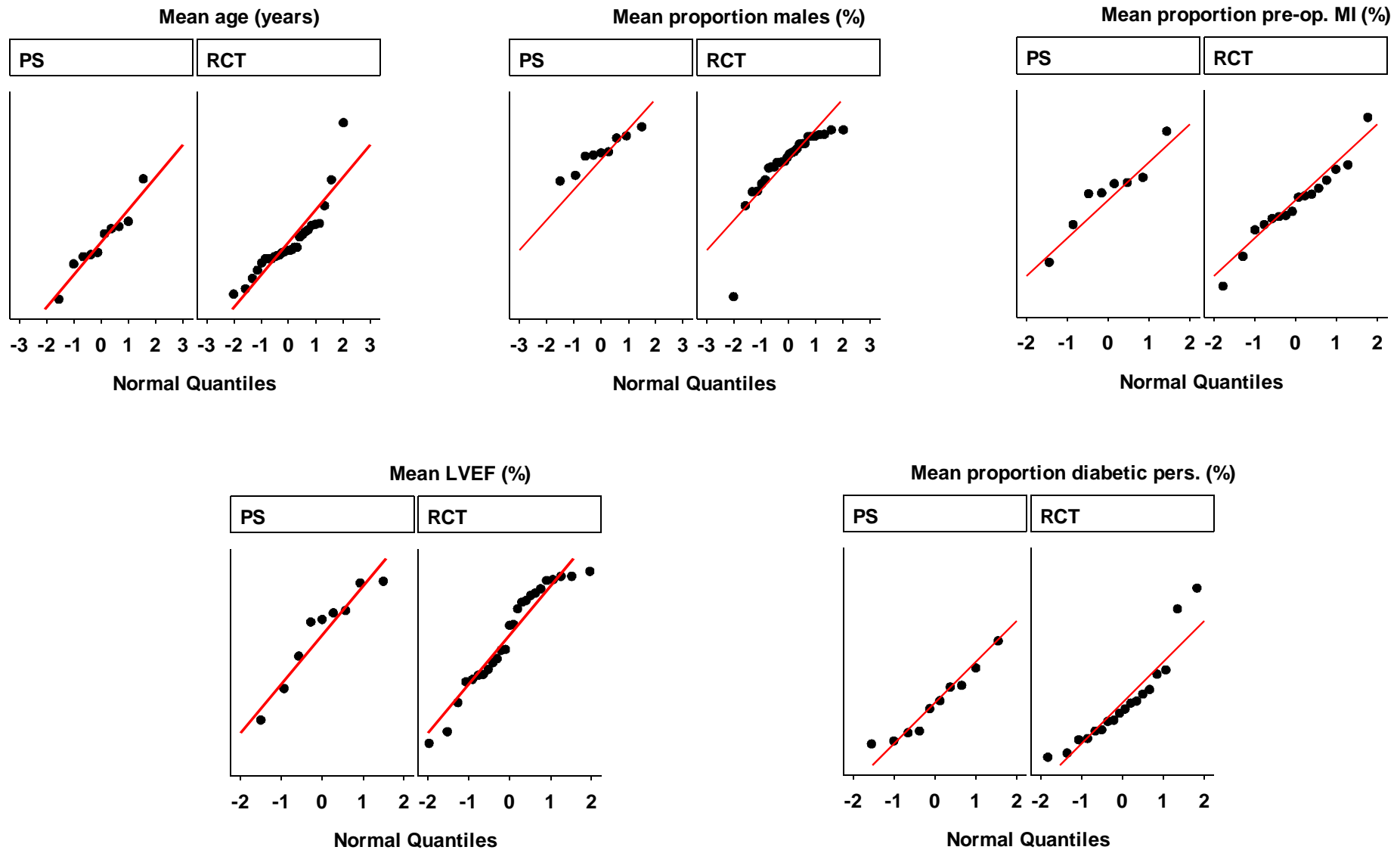
Ergebnisse II: Studien vor dem „Meta-Matching“

Meta-Confounder	PS-Analysen (N=28)	RCTs (N=51)	p-Wert	Standard. Diff. (%)
Studienregion			0.007	
Europa	17 (61%)	36 (71%)		
Nordamerika	10 (36%)	5 (10%)		
Sonstige	1 (3%)	10 (19%)		
Anzahl Zentren				
1	18 (65%)	47 (92%)	0.006	
>1	9 (32%)	3 (6%)		
Fehlend	1 (3%)	1 (2%)		
Mittl. Alter (Jahre)	65.8	63.1	0.002	75.1
Mittl. Anteil Männer (%)	72.1	77.1	0.138	-37.0
Mittl. Anteil prä-op. Herzinf.(%)	44.5	41.6	0.480	21.0
Mittl. LVEF (%)	58.8	62.7	0.033	-55.9
Mittl. Anteil Diabetiker (%)	26.2	24.4	0.595	13.9

Ergebnisse III: Studien *nach* dem „Meta-Matching“

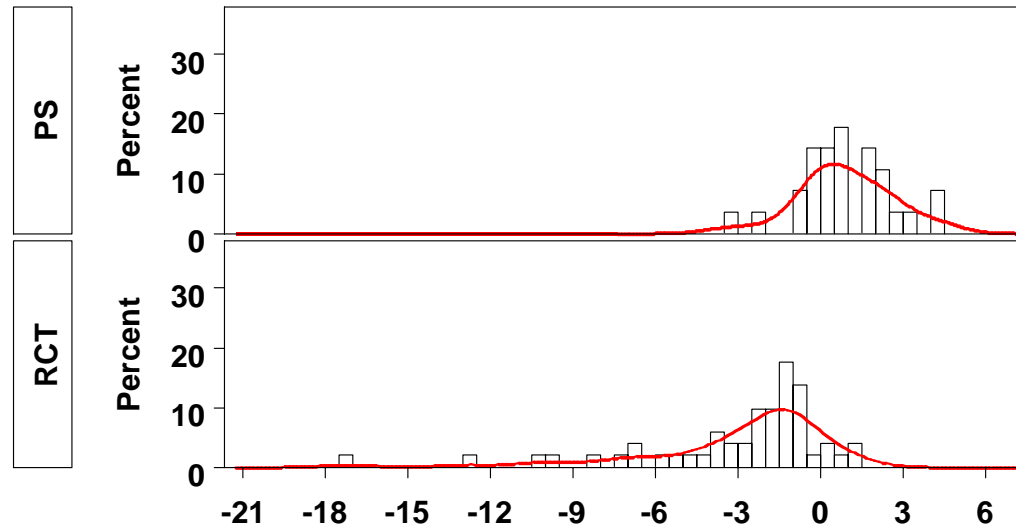
Meta-Confounder	PS-Analysen (N=10)	RCTs (N=29)	p-Wert	Standard. Diff. (%)
Studienregion			0.999	
Europa	8 (80%)	23 (80%)		
Nordamerika	1 (10%)	3 (10%)		
Sonstige	1 (10%)	3 (10%)		
Anzahl Zentren				
1	8 (80%)	25 (86%)	0.631	
>1	2 (20%)	3 (10%)		
Fehlend	0 (0%)	1 (4%)		
Mittl. Alter (Jahre)	64.1	63.9	0.916	3.9
Mittl. Anteil Männer (%)	80.5	76.9	0.431	30.5
Mittl. Anteil prä-op. Herzinf.(%)	44.0	39.9	0.530	27.6
Mittl. LVEF (%)	61.1	60.7	0.861	6.8
Mittl. Anteil Diabetiker (%)	24.8	25.2	0.925	-3.7

Ergebnisse IV: Verteilungen (Q-Q-Plots) der stetigen „Meta-Confounder“

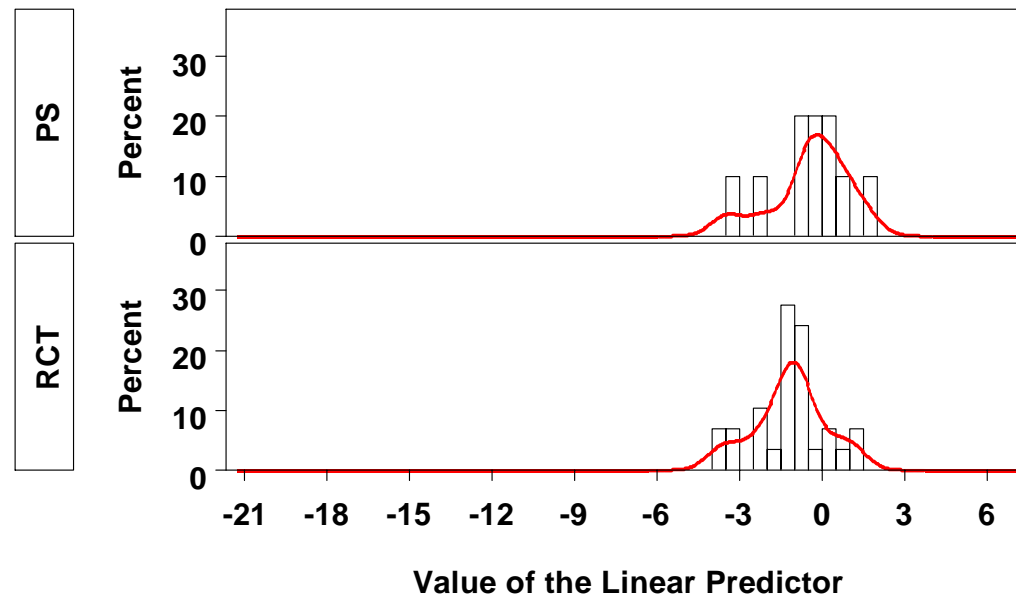


Ergebnisse V: Verteilung des linearen „Meta-Prädiktors“ vor und nach „Meta-Matching“

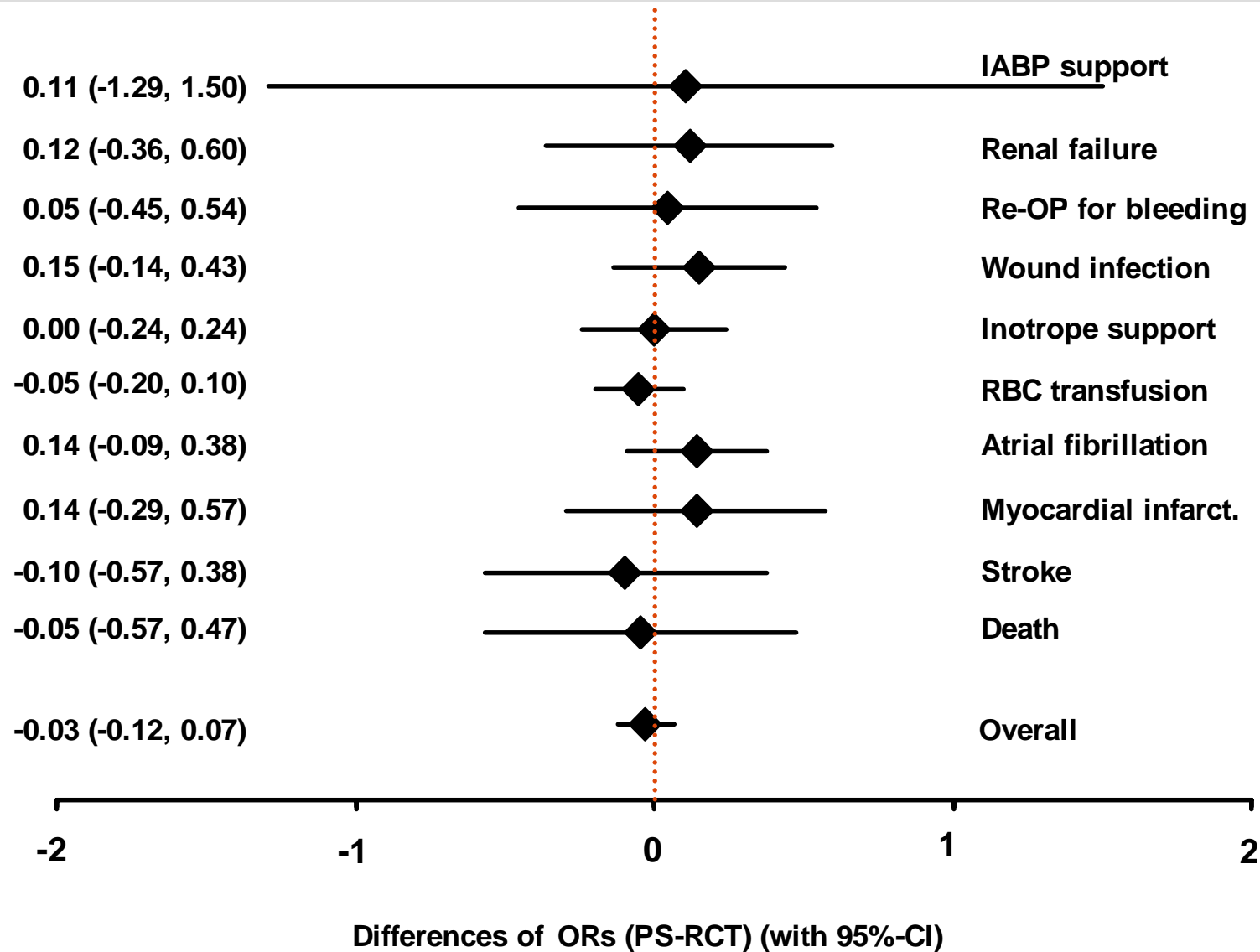
vorher



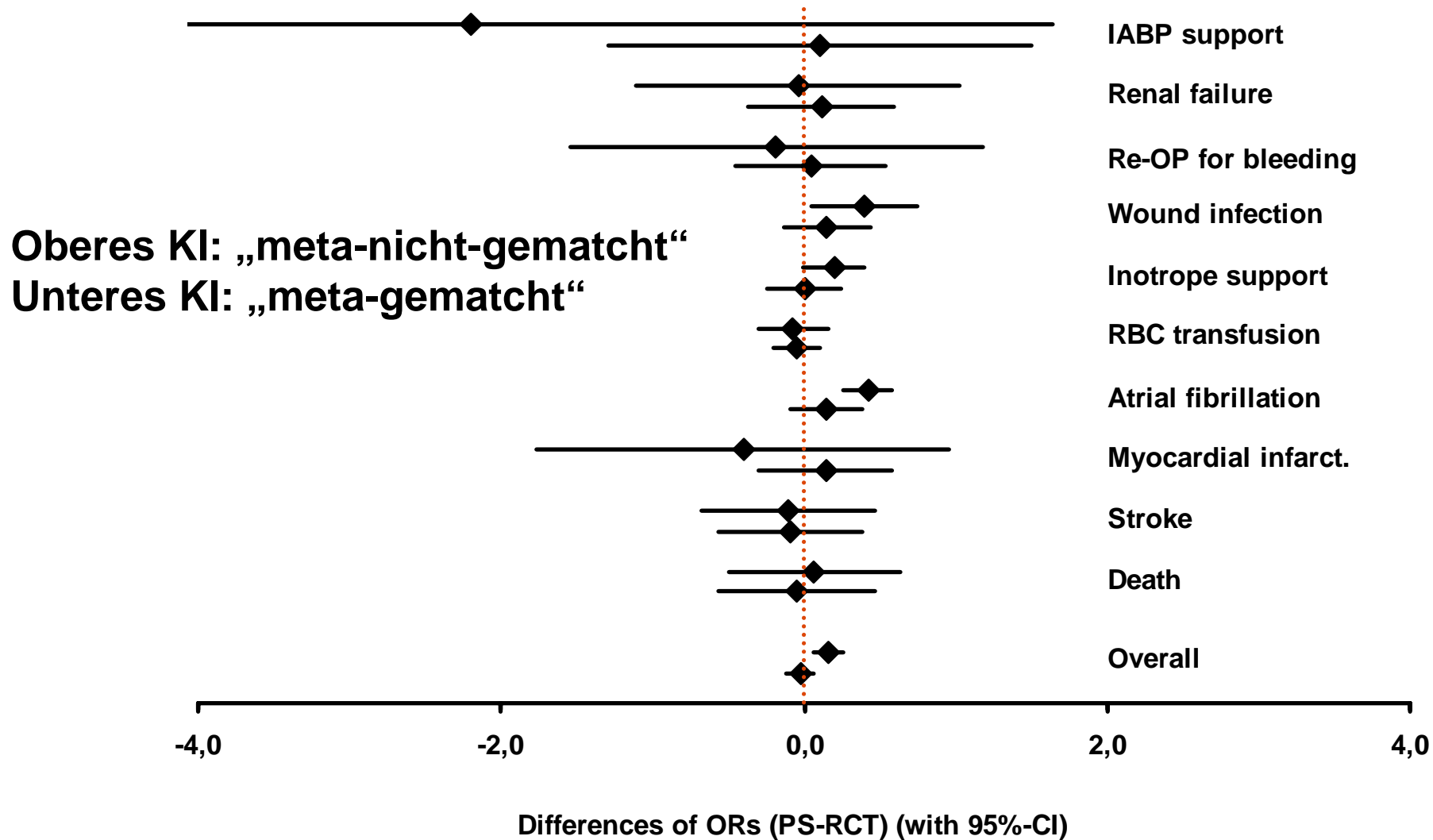
nachher



Ergebnisse VI: Differenzen der „Meta-ORs“ (PS-RCT) im „meta-gematchten“ Datensatz



Ergebnisse VII: Differenzen der „Meta-ORs“ (PS-RCT) im „meta-gematchten“ Datensatz und im „meta-nicht-gematchten“ Datensatz



Diskussion I

- In unserem Beispiel aus der Herzchirurgie waren die Unterschiede zwischen RCTs und PS-Analysen mit ähnlichen Populationen marginal
(Globale Diff. in Meta-ORs [95%-CI]: -0.027 [-0.119, 0.066])
→ **Kleiner Effekt der Randomisierung**
- Weitere **Vorteile** unserer Studie, die für eine gute Vergleichbarkeit von RCTs und Non-RCTs sorgen:
 - Identisches Design der Non-RCTs (PS)
 - Identisches Intervention und Kontrolle
 - Identische Zielgrößen in RCTs und Non-RCTs
 - Identische Follow-Up-Länge
 - Valide klinische Outcomes
 - RCTs und Non-RCTs in ähnlichen Zeiträumen durchgeführt

Diskussion II

- **Gefahren/Nachteile:**
 - Publication bias?
 - Vereinfachende Annahmen zu simpel?
 - „Meta-Residual Confounding?“
(Zur Erinnerung: Eigentlich brauchen wir eine „meta-randomisierte Studie“!)
 - Balanciertheit der „Meta-Confounder“ im „meta-gematchten“ Datensatz garantiert nicht die Balanciertheit für den einzelnen klinischen Outcome!

Diskussion III

- **In der Zukunft:**

Unsere Studie sollte unabhängig repliziert werden, am besten in einem anderen klinischen Bereich.

Auch nach einer Replikation werden RCTs nicht obsolet werden!

Allerdings: Die gegenwärtige Praxis, gut gemachte Non-RCTs (besser: PS-Analysen) aus systematischen Reviews für Therapieeffekte auszuschließen, sollte auf den Prüfstand!

Literatur

- 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.
- Kuss O, von Salviati B, Börgermann J. Off-pump versus On-Pump in Coronary Artery Bypass Grafting: A Systematic Review and Meta-Analysis of Propensity Score Analyses. In preparation.
- 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.
- 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 of Systematic Reviews* 2007, Issue 2.
- 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.
- Ioannidis JP, Haidich AB, Pappa M et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA*. 2001;286:821-830.