

Aus dem Institut für Medizinische Epidemiologie, Biometrie und Informatik der  
Medizinischen Fakultät der Martin-Luther-Universität-Halle-Wittenberg  
(Direktor: Prof. Dr. rer. nat. Johannes Haerting)

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

**Evidenz aus einer „Meta-Propensity-Score-Analyse“ in der Herzchirurgie**

Dissertation

zur Erlangung des akademischen Grades

Doktor der Medizin (Dr. med.)

vorgelegt

der Medizinischen Fakultät

der Martin-Luther-Universität Halle-Wittenberg

von Thomas Legler

geboren am 12.03.1981 in Pirna

Betreuer: PD Dr. sc. hum, rer. medic. habil. Oliver Kuß

Gutachter: 1. PD Dr. sc. hum, rer. medic. habil. Oliver Kuß  
2. Prof. Dr. med. Jan Gummert (Herz- und Diabeteszentrum NRW;  
Ruhr-Universität Bonn, Bad Oeynhausen)  
3. PD Dr. med. Stefan Sauerland (Institut f. Qualität und  
Wirtschaftlichkeit im Gesundheitswesen, Köln)

05.08.2010

17.05.2011

Meinen lieben Eltern,  
meiner lieben Freundin Rita.

## Referat

Randomisierte, kontrollierte, klinische Studien sind der Qualitätsstandard bei der Evaluation von Therapieeffekten. Die RCTs werden jedoch häufig wegen ihrer geringen externen Validität kritisiert <sup>[1]</sup>. Daneben gibt es Situationen in denen die Durchführung randomisierter Studien „unnötig, ungeeignet, unmöglich oder ungenügend“ ist <sup>[2]</sup>. Hier stellen nicht randomisierte Beobachtungsstudien eine attraktive Alternative dar. Aufgrund der fehlenden Randomisation kommt es durch Confounding zu einer Verzerrung der Effektschätzer. Zur Vermeidung bzw. Verringerung dieser Verzerrung ist eine Confounderadjustierung notwendig. Eine sehr valide und mit zunehmender Beliebtheit verwendete Methode dafür ist die Propensity-Score-Analyse <sup>[3]</sup>. In dieser Arbeit werden die Outcomes von RCTs und PS-Analysen mit einander verglichen. Für diesen Vergleich wurde im Rahmen dieser Arbeit eine „Meta-Propensity-Score-Analyse“ durchgeführt, um ein „Meta-Confounding“ zu vermeiden. Die zugrunde liegende klinische Fragestellung war der Vergleich der Off- und On-Pump-Technik bei Bypassoperation in der Kardiochirurgie. Dazu wurden in einem systematischen Review mit den üblichen Anforderungen (reproduzierbare Suche, zwei verblindete Reviewer (OK, TL), Konsensusfindung mit einem dritten Reviewer (JB), strukturierte Datenerfassung, a priori definierte Ein- und Ausschlusskriterien für Studien, klinische Zielgrößen, „Meta-Confounder“ und „Meta-Zielgrößen“) alle randomisierten Studien und PS-Analysen zu dieser Fragestellung gesammelt und die zugrunde liegenden Studieneigenschaften und Patientenpopulationen mithilfe eines Meta-Propensity-Score-Modells „meta-gematcht“. Es gingen 51 RCTs und 28 PS-Analysen in die Studie ein. Es ergaben sich 7 Meta-Confounder, die in den Matching-Prozess eingingen. Im „meta-gematchten“ Datensatz finden sich Daten von 29 RCTs und 10 PS-Analysen. Es konnte kein signifikanter Unterschied zwischen den Ergebnissen aus RCTs und PS-Analysen ( $OR_{PS-RCT_{gesamt}} [-0.03; -0.12 - 0.07]$ ) eruiert werden. Ohne Meta-Matching-Prozess konnte hingegen ein signifikanter Unterschied ( $OR_{PS-RCT_{gesamt}} [0.17; 0.07 - 0.26]$ ) gargestellt werden. Sowohl mittels RCTs ( $OR_{gesamt} [0.6132; 0.5368 - 0.7005]$ ) als auch PS-Analysen ( $OR_{gesamt} [0.5867; 0.5448 - 0.6318]$ ) konnte ein signifikant besserer Therapieeffekt auf Seiten der Off-Pump-Technik verzeichnet werden.

Legler, Thomas: Gibt es einen Unterschied zwischen randomisierten und nicht randomisierten Studien? Evidenz aus einer „Meta-Propensity-Score-Analyse“ in der Herzchirurgie. Halle/Saale, Univ., Med. Fak., Diss., 69 Seiten, 2010

## Inhaltsverzeichnis

1.	Einleitung .....	1
1.1	Hintergrund der vorliegenden Arbeit.....	1
1.2	Randomisierte, kontrollierte, klinische Studien .....	2
1.2.1	Methodische Vorteile.....	2
1.2.2	Methodische Nachteile .....	3
1.3	Nicht randomisierte Beobachtungsstudien .....	6
1.3.1	Methodische Vorteile.....	6
1.3.2	Methodische Nachteile .....	6
1.4	Bias und Confounding .....	7
1.4.1	Confounderadjustierung .....	9
1.5	Propensity score .....	12
1.5.1	Begriffserklärung .....	12
1.5.2	Anwendung .....	13
1.5.3	Methodische Vorteile.....	15
1.6	Gegenstand der vorliegenden Arbeit.....	16
1.6.1	Problematik des Vergleichs .....	16
1.7	Aortokoronare Bypassoperation .....	20
1.7.1	Zugangsweg .....	20
1.7.2	On-Pump-Technik .....	20
1.7.3	Off-Pump-Technik .....	22
2.	Zielsetzung der vorliegenden Arbeit .....	24
2.1	Fragestellung .....	24
2.2	Hypothesen.....	24
3.	Material und Methodik .....	25
3.1	Studienthemen.....	25
3.2	Studiensuche und Datenextraktion aus RCT .....	25
3.2.1	Studiensuche .....	25
3.2.2	Datenextraktion .....	25
3.2.3	Studieneinschluss .....	27
3.3	Datensuche und Datenextraktion aus PS-Analysen .....	27
3.3.1	Studiensuche .....	27
3.3.2	Datenextraktion .....	28
3.3.3	Studieneinschluss .....	28
3.4	Statistische Methoden .....	29
3.4.1	Datenbeschreibung .....	29
3.4.2	Umgang mit unterschiedlichen Ausdrücken der Verteilungsbeschreibung.....	29
3.4.3	Umgang mit fehlenden Daten.....	30
3.4.4	Meta-Propensity-score-Modell.....	31
3.4.5	Klinische Zielgrößen.....	31
4.	Ergebnisse .....	32
4.1	Studieneinschluss .....	32
4.1.1	Einschluss der randomisierten, kontrollierten klinischen Studien.....	32
4.1.2	Einschluss der Propensity score Analysen .....	34
4.2	Verteilung der Meta-Confounder vor Meta-Matching .....	36
4.3	Verteilung der Meta-Confounder nach Meta-Matching .....	38
4.4	Verteilung der Meta-Confounder zwischen den ungematchten Studien..	39
4.5	Klinische Zielgrößen.....	41
4.5.1	Effektschätzer der klinischen Zielgrößen.....	41
4.5.2	Differenzen der Meta-Odds-Ratios .....	43
5.	Diskussion.....	46

## Inhaltsverzeichnis

---

6.	Zusammenfassung.....	51
7.	Literaturverzeichnis .....	52
8.	Anhang .....	59
8.1	Auswertungsbogen .....	59
8.2	Eingeschlossene Studien .....	64
8.3	Ausgeschlossenen Studien .....	66
9.	Thesen.....	68

## Verzeichnis der Abkürzungen und Symbole

°C	Grad Celsius
a	Jahre = Lebensjahre
ACVB	Aortocoronarer Venenbypass
BMI	body mass index
BvS	Benita von Salviati
bzw.	beziehungsweise
CABG	coronary artery bypass grafting = CCAB
CCAB	conventionally coronary artery bypass = CABG
CCTR	Cochrane Central Register of Controlled Trials
CDSR	Cochrane Database of Systematic Reviews
CI	Konfidenzintervall
CO <sub>2</sub>	Kohlenstoffdioxid
DARE	Database of Abstract of Reviews of Effects
DP	Doppelpublikation
EBM	Evidence based Medicine
EK	Erythrozytenkonzentrat
EKs	Erythrozytenkonzentrate
etc.	et cetera
geg.	gegeben
HLM	Herzlungenmaschine
IABP	intraaortic ballon pump
ITT	Intention to treat
J	Jahre = Lebensjahre
kg	Kilogramm
KG	Körpergewicht
kont. MC	Kontinuierliche Meta-Confounder
LAD	left anterior descending = RIVA
LIMA	left internal mammarian artery = Arteria. thoracica interna sinistra
LVEF	linksventrikuläre Ejektionsfraktion
m	matched
männl.	männlich
MI	Myokardinfarkt
MIDCAB	minimally invasive direct coronary artery bypass
ml	Milliliter

## Abkürzungsverzeichnis

---

Mm	Millimeter
N	Anzahl
NIS	Keine Information zur Studienpopulation
nm	non matched
NO	fehlende Angaben zu binären Outcomes
NR	fehlende Randomisation
nRCT	non randomized controlled trial = nicht randomisierte, kontrollierte, klinische Studie
nRCTs	non randomized controlled trials
o. g.	oben genannt
OK	Dr. Oliver Kuß
OPCAB	off-pump coronary artery bypass grafting
OR	Odds Ratio
OT	andere Therapien
postop.	postoperativ
präop.	präoperativ
PS	Propensity Score
PS-Analyse	Propensity-Score-Analyse
RCT	randomized controlled trial = randomisierte, kontrollierte, klinische Studie
RCTs	randomized controlled trials
RIVA	Ramus interventricularis anterior
s. u.	siehe unten
SD	standardisierte Differenz
SR	systematischer Review
TL	Thomas Legler
usw.	und so weiter
uvm.	und viele mehr
vgl.	vergleiche
z.B.	zum Beispiel

# 1. Einleitung

## 1.1 Hintergrund der vorliegenden Arbeit

Als „Gold-Standard“ für die Evaluation von Therapieeffekten werden im Allgemeinen die randomisierten, kontrollierten, klinischen Studien angesehen<sup>[4; 5; 6; 7; 8]</sup>. Als Metaanalysen zusammengefasst, stehen diese unter anderem laut der Ärztlichen Zentralstelle für Qualitätssicherung auf Platz Eins in der Evidenzhierarchie, gefolgt von nicht randomisierten, kontrollierten Beobachtungsstudien, klinischen Berichten und Expertenmeinungen<sup>[7; 9; 10]</sup>. Die Ergebnisse aus RCTs bilden die valideste Grundlage für die Wirksamkeitsbewertung medizinischer Maßnahmen<sup>[26]</sup>. Dabei ist zu beachten, dass zwischen externer<sup>1</sup> und interner<sup>2</sup> Validität unterschieden wird<sup>[12; 13]</sup>. Vor allem hinsichtlich der Aussagekraft gegenüber der Stichproben<sup>3</sup>- und Gesamtpopulation<sup>4</sup> spielt diese Unterteilung bei dem Vergleich von Therapiemethoden eine entscheidende Rolle. Die Kenntnis über die innere und äußere Validität einer Studienform ist somit von hoher Relevanz.

Theoretisch gibt es zwischen den einzelnen Studienformen vor allem in Bezug auf die Validität gravierende Differenzen. Im Folgenden soll auf diese Unterschiede, die Grundmerkmale sowie die methodischen Vor- und Nachteile der randomisierten und nicht-randomisierten, kontrollierten, klinischen Studien eingegangen werden.

---

<sup>1</sup> Die externe Validität wird als die äußere Gültigkeit einer Studie verstanden und beschreibt die Verallgemeinerungsfähigkeit der Studienergebnisse auf die Gesamtpopulation<sup>[13]</sup>.

<sup>2</sup> Die interne Validität wird als innere Gültigkeit einer Studie verstanden und beschreibt die Plausibilität bzw. Richtigkeit der erhaltenen Ergebnisse. Eine hohe interne Validität führt zu einem hohen Maß an Reproduzierbarkeit der erhaltenen Ergebnisse bei der Wiederholung einer Studie unter gleichen Bedingungen<sup>[13]</sup>.

<sup>3</sup> Die Stichprobenpopulation stellt einen Teil der Gesamtpopulation dar. An diesem Anteil wird der Therapieeffekt im Rahmen einer Studie untersucht<sup>[13]</sup>.

<sup>4</sup> Die Gesamtpopulation beschreibt den Zusammenschluss aller an derselben Erkrankung Leidender. Der Therapieeffekt wird hingegen an einer Stichprobenpopulation untersucht<sup>[13]</sup>.

### 1.2 Randomisierte, kontrollierte, klinische Studien

#### 1.2.1 Methodische Vorteile

1948 gilt anlässlich der in diesem Jahr erstmalig publizierten Anwendung als das Geburtsjahr der randomisierten, kontrollierten, klinischen Studien [7; 10; 11; 14]. Der große Vorteil dieser Studienform liegt in der zufälligen Verteilung der Patienten auf die Untersuchungs- und Kontrollgruppe, in Analogie zum Werfen einer Münze [13]. Entsprechend werden sowohl die bekannten als auch unbekannt Merkmale der an der Studie teilnehmende Probanden gleichmäßig auf die Beobachtungs- und die Kontrollgruppe verteilt. Die Patienteneigenschaften können einen entscheidenden Einfluss auf die Studienergebnisse nehmen. Deshalb ist die zufällige und homogene Aufteilung dieser Merkmale ein bedeutendes Argument, warum randomisierte, kontrollierte, klinische Studien auch als experimentelle Studien bezeichnet werden [10; 12]. Es werden Bedingungen geschaffen, ähnlich einem Experiment im klassischen Sinne, die die Reproduzierbarkeit der Ergebnisse maximieren.

Die sorgfältige Auswahl und Festlegung der Einschluss- bzw. Ausschlusskriterien sorgt zusätzlich für eine Verringerung der Heterogenität der Gruppen und somit zu einer Erhöhung der internen Validität.

Des Weiteren unterstützt, wenn praktiziert, eine Verblindung des Untersuchenden, der Untersuchten und des für die Datenauswertung Zuständigen die Minimierung des so genannten Hawthorne-Effektes<sup>5</sup> und somit der unbewussten aber auch bewussten Beeinflussung der Outcomes durch Beobachtete und Beobachter [13].

---

<sup>5</sup> „Der Hawthorne-Effekt ist ein Phänomen, bei dem Menschen dazu neigen, ihr Verhalten zu verändern, wenn sie Ziel eines besonderen Interesses und besonderer Aufmerksamkeit in einer Studie sind, unabhängig von der Natur der therapeutischen Intervention, der sie zuteil werden könnten. Die Ursachen dieses veränderten Verhaltens sind nicht klar. Die Patienten sind bestrebt, ihre Ärzte zufrieden zu stellen und ihnen das Gefühl zu vermitteln, erfolgreich zu sein. Ebenso wollen Patienten, die freiwillig an einer Studie teilnehmen, ihren Beitrag dazu leisten, dass „gute“ Ergebnisse erzielt werden“. (Fletcher RH, Fletcher SW, Wagner EH: Klinische Epidemiologie, Grundlagen und Anwendung. Deutschsprachige Ausg. Adaptiert und hrsg. von Johannes Haerting und Christoph Rink. Ullstein Medical, Wiesbaden, 1999, S.199.)

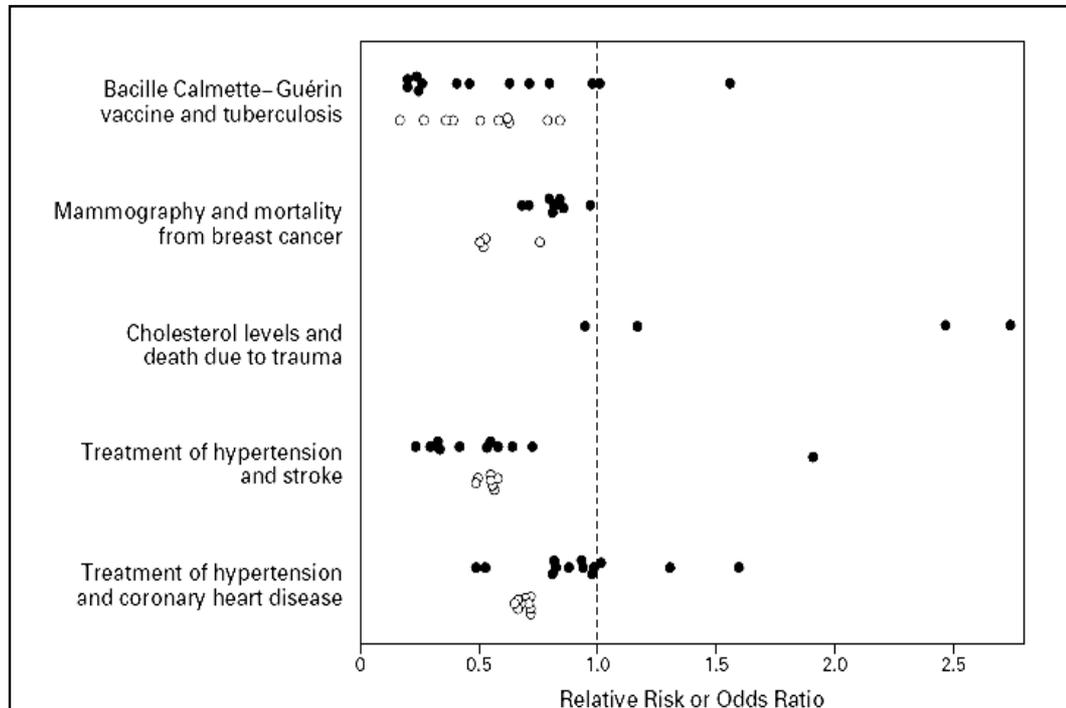
### 1.2.2 Methodische Nachteile

Die Vorteile randomisierter, kontrollierter, klinischer Studien sind gleichzeitig auch Ursache für die limitierte Anwendung dieser Studienform. Die strengen Einschlusskriterien bei der Stichprobenauswahl, die notwendig sind, um ein Minimum an Heterogenität der zu vergleichenden Gruppen zu erreichen und unzuverlässige Patienten schon im Vorfeld auszuschließen, erzeugen ein zwar homogenes, aber auf Kosten der externen Validität ein hochselektiertes Patientengut <sup>[10; 15]</sup>. Zusätzlich sind diese Studienteilnehmer größtenteils jünger und gesünder als der Durchschnittspatient <sup>[12; 16; 17; 18]</sup>.

Die Durchführung einer randomisierten, kontrollierten, klinischen Studie ist in der Regel mit enorm hohen Kosten verbunden <sup>[13]</sup>. Diese lassen in Verbindung mit den sehr stringenten Selektionskriterien nur eine geringe Populationsgröße zu, was wiederum eine kürzere Dauer der Studie und der Verlaufsbeobachtung und folglich eine geringere Abstraktionsfähigkeit, speziell in Anbetracht chronisch Kranker, zur Folge hat. Außerdem wirken sich kleine Patientengruppen sehr nachteilig auf die Untersuchung und den Vergleich relativ ungewöhnlicher oder sehr seltener Outcomes aus. Entsprechend einer Faustregel müssen 3 x n Personen beobachtet werden, um eine gute Chance zu haben, ein 1/n-Outcome nachweisen und somit die zu untersuchende Therapien in Bezug auf dieses Outcome vergleichen zu können <sup>[19]</sup>. Unter Bezugnahme dieser Punkte ist es nachvollziehbar, dass randomisierte, kontrollierte Studien meist an akademischen oder wissenschaftlich aktiven medizinischen Zentren durchgeführt werden und sehr viel seltener in kleineren Kliniken oder ambulanten Versorgungseinrichtungen, welche vor allem in außerstädtischen Gebieten die Hauptanlaufpunkte der Durchschnittspatienten darstellen <sup>[10]</sup>.

Concato et al. <sup>[7]</sup> konnten durch einen Vergleich von randomisierten und nicht randomisierten Studien anhand fünf klinischer Fragestellungen eine Auswirkung kleiner Studienpopulationen auf die Studienergebnisse aufdecken. Es zeigte sich, dass die Ergebnisse mehrerer RCTs einer hohen Streuung unterliegen (Abbildung 1). Aus diesem Grund sollten Schlussfolgerungen über die Wirkungen bestimmter Behandlungsformen nicht aus einzelnen Studien, sondern aus Metaanalysen gezogen werden. Die geringere Anzahl an untersuchten Probanden hat neben der o. g. hohen Streuung der einzelnen Ergebnisse auch

eine geringere statistische Trennschärfe<sup>6</sup> einer Studie zur Folge<sup>[13]</sup>. Grund dafür ist das erhöhte Risiko für das Auftreten eines  $\beta$ -Fehlers<sup>7</sup>. Folglich ist die Chance, einen tatsächlich vorhandenen Therapieeffekt mittels einer relativ kleinen Studienpopulation zu finden, gering.



**Abbildung 1: Odds Ratio aus dem Vergleich randomisierter (schwarze Punkte) und nicht-randomisierter Studien (transparente Punkte) [7]**

Einen weiteren Nachteil randomisierter, kontrollierter, klinischer Studien stellt der potenzielle Effekt von systematischer Patientenaufklärung auf den Therapieverlauf und die Outcomes dar. Ohne Einwilligungserklärung wäre eine derartige Studie nicht durchzuführen. Kaptchuk et al.<sup>[20]</sup> untersuchten diesen Effekt in einer randomisierten, doppelblinden, Placebo-kontrollierten Studie. Zum Vergleich standen sich eine Gruppe mit Patientenaufklärung und eine Gruppe ohne Aufklärung gegenüber. Es wurde die analgetische Wirkung von Naproxen verglichen mit der eines Placebos untersucht. Patienten, die prätherapeutisch aufgeklärt wurden und ihr Einverständnis gaben, berichteten sowohl unter Naproxen, als auch unter Placebo über eine effektivere Schmerzlinderung als die Patienten ohne vorherige Aufklärung und Zustimmung. Des Weiteren war der

<sup>6</sup> Die Statistische Trennschärfe ( $1 - \text{Wahrscheinlichkeit des } \beta\text{-Fehlers (1-p}_\beta)$ ) gibt die Wahrscheinlichkeit eines statistischen Tests an, bei wirklich vorhandener Differenz, einen signifikanten Unterschied zu eruieren<sup>[13]</sup>.

<sup>7</sup> Der  $\beta$ -Fehler gibt die Wahrscheinlichkeit des Zustandekommens eines falschnegativen Ergebnisses, die Arbeitshypothese betreffend, an<sup>[13]</sup>.

Ergebnisunterschied zwischen der Naproxen- und Placebogruppe in Anbetracht des Therapieeffektes mit Patientenaufklärung geringer als in der Vergleichsgruppe. Es lässt sich folglich nicht ableiten ob ein erhaltener Therapieeffekt tatsächlich besteht oder eher Folge der notwendigen Patientenaufklärung und Einwilligung zur Teilnahme an einer RCT ist.

Weiterhin ist festzuhalten, dass es Situationen gibt, in denen die Durchführung einer RCT unnötig, ungeeignet, unmöglich, unzureichend oder schlicht unethisch ist [2; 5; 9]. Untersucht man zwei Therapiemethoden und der zu erwartende Unterschied in den gemessenen Therapieeffekten wäre minimal, so wäre der Einsatz einer RCT in Bezug auf die Kosten eher unnötig. Weiter ließe die kleine Studienpopulation aufgrund des erhöhten Risikos für das Auftreten eines  $\beta$ -Fehlers eine nur geringe Trennschärfe zu. Die Durchführung einer RCT wäre bei der zu erwartenden geringen Ergebnisdifferenz somit ungeeignet und unzureichend. Schließlich wäre es schlicht unethisch, einem Patienten eine bewährte Behandlung mit wissenschaftlich gesichertem und klinisch erprobten Therapieeffekt vorzuenthalten, um einen Vergleich mit einer neueren, nicht sicher besseren Methode durchführen zu können.

In diesen Situationen käme die Umsetzung einer nicht randomisierten Beobachtungsstudie im Sinne von Kohortenstudien in Frage.

### 1.3 Nicht randomisierte Beobachtungsstudien

#### 1.3.1 Methodische Vorteile

Die nicht randomisierten Beobachtungsstudien werden vornehmlich zur Ermittlung von Risiko- und Prognosefaktoren, aber auch als attraktive Alternative zur Evaluation von Therapieeffekten eingesetzt, wenn RCTs beispielsweise aus oben genannten Gründen nicht möglich sind. Die bedeutendsten Vorzüge sind die relativ einfache Realisierbarkeit, die geringeren Kosten und die Möglichkeit der größeren Studienpopulationen mit höherer Merkmalsvielfalt und folglich die höhere externe Validität <sup>[9; 13]</sup>. Es besteht aufgrund der größeren Probandenanzahl eine höhere Chance einen geringen aber tatsächlich vorhandenen Therapieeffekt darstellen zu können. Gründe dafür sind die höhere statistische Trennschärfe und die höhere Chance, gering ausfallende Outcomes festzustellen und somit vergleichen zu können (vgl. Kapitel 1.2.2).

Zusätzlich von Vorteil ist die Möglichkeit, retro<sup>8</sup>- oder prospektiv<sup>9</sup> beobachten zu können, was eine wichtige Funktion in Hinsicht auf die Fülle an vorhandenen Daten in den Kliniken und ambulanten Zentren ist. Ein weiterer positiver Effekt, vor allem bezüglich einer größeren Studienpopulation, ist die Möglichkeit, Vergleiche erarbeiten zu können, bei denen Daten aus begleitenden und Daten aus historischen Beobachtungen genutzt werden <sup>[13]</sup>.

#### 1.3.2 Methodische Nachteile

Prinzipiell nachteilig ist die Art der Patientenverteilung auf den Therapie- und Kontrollarm. Diese erfolgt bei prospektiven Beobachtungen aus klinischen Gesichtspunkten, dem Ermessen des zuständigen Arztes. Diese Art der Patientenselektion und -verteilung ist verantwortlich für die zumeist geringere interne Validität nicht randomisierter Beobachtungsstudien. Die gegenüberstehenden Gruppen sind in Bezug auf ihre Eigenschaften (Patientenalter, Geschlecht, Komorbidität, uvm.) mit großer Wahrscheinlichkeit sehr unterschiedlich. Für den Leser ist es nur schwer ersichtlich, ob ein positiver, negativer oder gleicher Behandlungseffekt Folge der untersuchten Intervention oder der Einflussnahme durch die unterschiedlichen Patientenmerkmale ist. Zum anderen ist bei historischen Beobachtungen eine Verteilung der Patienten auf

---

<sup>8</sup> historische Kohortenstudie

<sup>9</sup> begleitende Kohortenstudie

Therapie- und Kontrollgruppe bereits abgeschlossen, ohne Beachtung einer eventuellen, späteren, vergleichenden Untersuchung von Behandlungseffekten.

Einige Autoren gehen mit ihren Ratschlägen sogar soweit, dass sie dem Leser empfehlen, Publikationen nicht randomisierter Studien beiseite zu legen um sich einem Artikel, basierend auf einer RCT, zu widmen<sup>[9; 21]</sup>.

### **1.4 Bias<sup>[13]</sup> und Confounding<sup>[5; 13]</sup>**

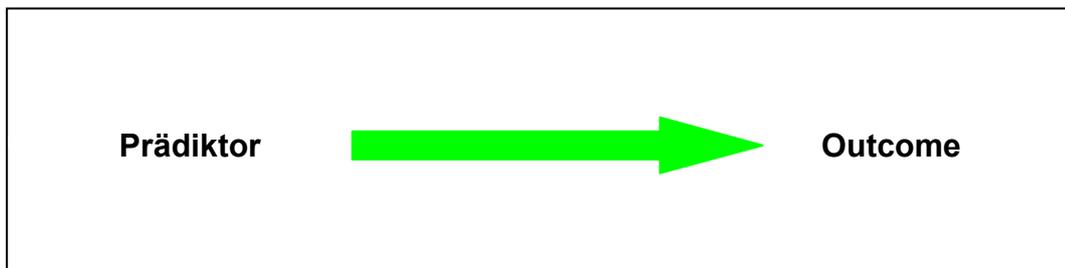
Bei einer Gegenüberstellung zweier Behandlungsstrategien ist lediglich ein Unterschied in den Interventionen und den Outcomes legitim bzw. erwünscht. Zusätzliche Differenzen zwischen den Determinanten eines klinischen Zustandes, beispielsweise den Patientenmerkmalen oder Behandlungs- und Beobachtungsstrategien beider Gruppen, führen zwangsläufig zu Verzerrungen. Die erhaltenen Outcomes neigen daraufhin dazu, von den tatsächlichen Werten abzuweichen. Infolgedessen ist eine Reproduzierbarkeit und ordnungsgemäße Interpretation der Outcomes nicht erreichbar.

Bias<sup>10</sup> „kann auf jeder Stufe des medizinischen Erkenntnisprozesses auftreten“ [Fletcher RH, Fletcher SW, Wagner EH: Klinische Epidemiologie, Grundlagen und Anwendung. Deutschsprachige Ausg. Adaptiert und hrsg. von Johannes Haerting und Christoph Rink. Ullstein Medical, Wiesbaden, 1999, S. 10]. In Anbetracht der technischen Umstände der jeweiligen Studienform und der Studiendurchführung unterscheidet man Selektions- und Informationsbias. Systematische Fehler, die während des Rekrutierungsprozesses der Patienten auftreten, gehören in die Gruppe der Selektionsbias. Die Probanden könnten von Haus aus ein höheres bzw. geringeres Basisrisiko für eine Erkrankung und deren Ausgang haben. Besteht der Fall, dass Praktiken zur Informationsgewinnung nicht uniform bzw. fehlerhaft sind, spricht man von Informationsbias. Hierbei führen nicht genormte, uneinheitliche Messmethoden oder fehlerhafte und unzureichende Dokumentation zu einem nicht validen Vergleich und infolgedessen zu einer möglicherweise irreführenden Schlussfolgerung. Deshalb müssen Selektions- und Informationsbias bereits in der Planungsphase einer Studie vermieden werden.

---

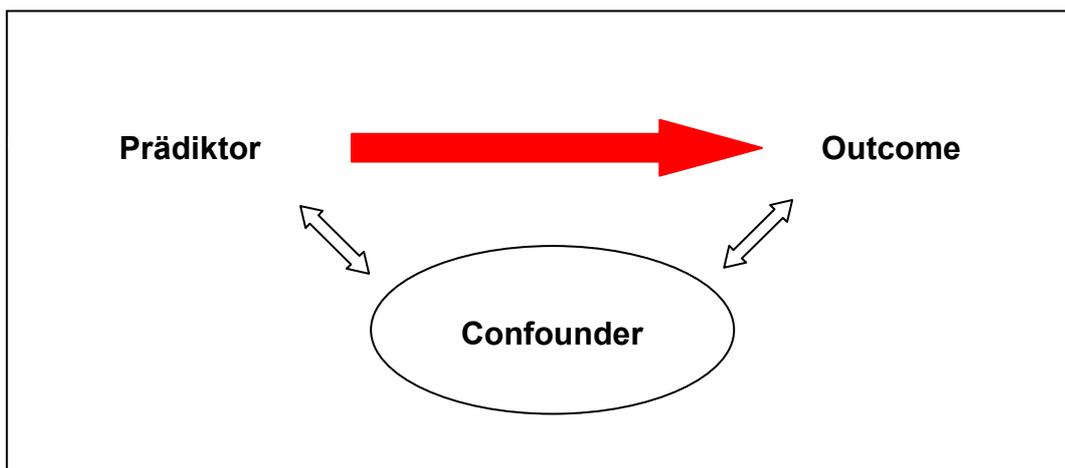
<sup>10</sup> Bias beschreibt eine Verzerrung der Ergebnisse infolge eines systematischen Fehlers.

Besteht ein direkter statistischer Zusammenhang zwischen dem für den Vergleich primär irrelevanten Faktor (Confounder), beispielsweise dem Geschlecht, der Komorbidität, eventuellen Kointerventionen, etc. und dem gemessen Effekt (Outcome), ist davon auszugehen, dass dieser Faktor (Confounder) den Therapieeffekt (Outcome) bzw. die Beziehung zwischen Prädiktor (Therapie) und Outcome beeinflusst. Es wäre also nicht möglich, auf eine konkrete Prädiktor<sup>11</sup>-Outcome<sup>12</sup>-Beziehung, wie in Abbildung 2, zu schließen. Das heißt jedoch nicht, dass die o. g. Basismerkmale (Confounder) der an einer klinischen Beobachtung teilnehmenden Patienten unerheblich für die Schlussfolgerung und Ergebnisbewertung sind. Sie sind aber ein verzerrendes Moment in dem reinen Zusammenhang zwischen Prädiktor und Outcome.



**Abbildung 2: Beziehung zwischen Prädiktor und Outcome ohne Confounding**

Tritt das eben Beschriebene ein, spricht man von Confounding. Dieser Zusammenhang ist in Abbildung 3 in Form eines Confunderdreiecks dargestellt.



**Abbildung 3: Beziehung zwischen Prädiktor und Outcome mit Confounding**  
[5; 13]

<sup>11</sup> Der Prädiktor beschreibt die zu untersuchende Behandlungsform, z.B. im Sinne einer pharmakologischen, interventionellen oder operativen Therapie <sup>[5]</sup>.

<sup>12</sup> Das Outcome beschreibt das Ergebnis/Resultat <sup>[5]</sup>.

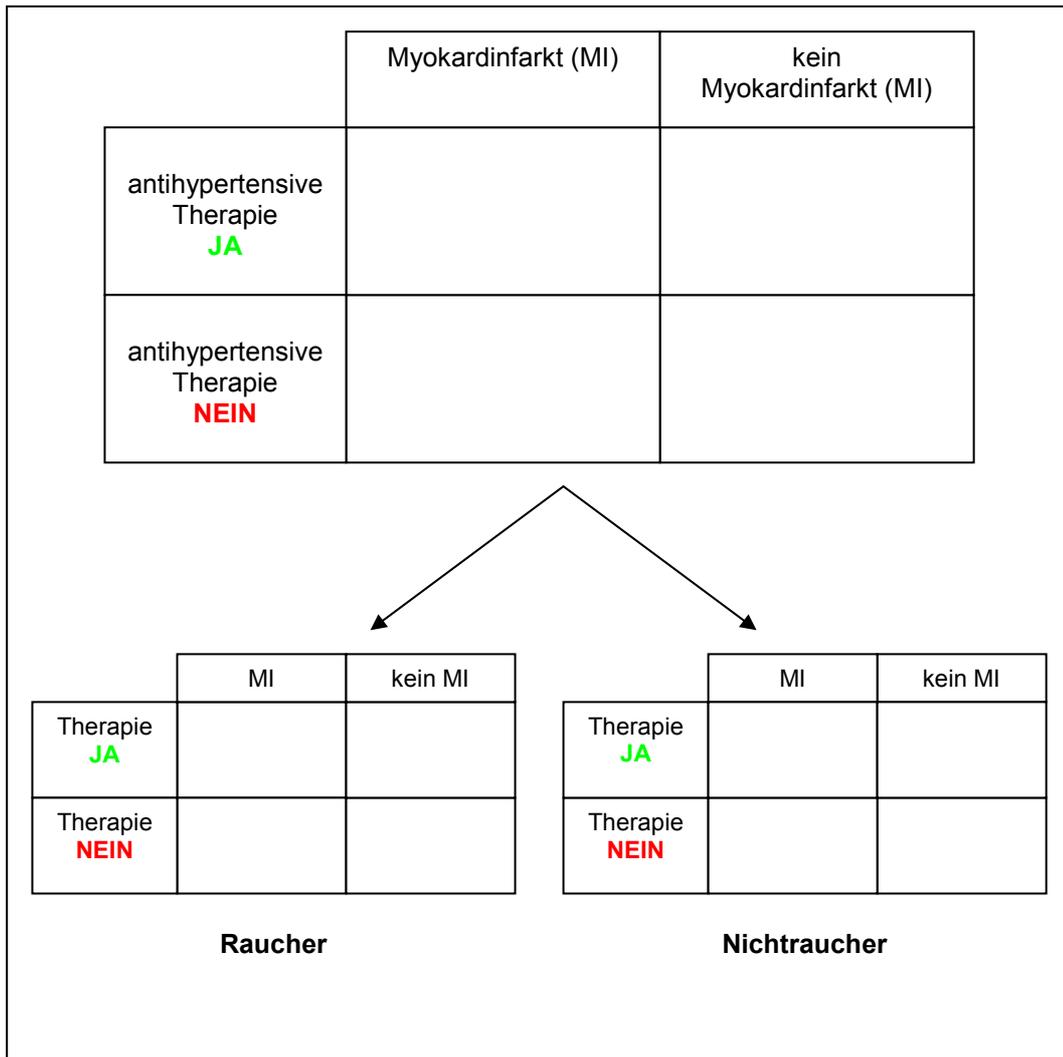
Um eine Gleichheit bezüglich des Basisrisikos für eine Erkrankung zu erzielen, ist es anzustreben, Patientenmerkmale möglichst gleich auf die Patientengruppen (Therapiegruppe/experimentelle Gruppe und Kontrollgruppe/Placebogruppe) zu verteilen oder die Beziehung zwischen der zu evaluierenden Therapie und dem Ergebnis für einen oder mehrere Confounder zu adjustieren.

### 1.4.1 Confounderadjustierung <sup>[13]</sup>

Die effektivste und einfachste Art, alle möglichen Confounder gleich auf die Therapie- und Kontrollgruppe zu verteilen, ist die Randomisation. Dadurch wird bereits in der Rekrutierungsphase einer Studie das Auftreten von Confounding verhindert bzw. minimiert. Bei der Durchführung nicht randomisierter Beobachtungsstudien hingegen kann eine Confounderadjustierung oder -kontrolle nur in der Phase der Datenanalyse erfolgen.

Die einfachste sowie eine der üblichsten und aufschlussreichsten Arten, die Zusammenhänge zwischen dem einzelnen Confounder und dem Studienresultat darzulegen, ist die Stratifizierung. Es wird jeweils einer möglicherweise verzerrenden Variable (potenzieller Confounder) bezüglich des zu untersuchenden Outcomes eine Vierfelder- bzw. mehrerer verzerrender Variablen eine Kontingenztafel zugeordnet. Daraus wird ersichtlich, ob es wirklich eine Beeinflussung des Ergebnisses durch den möglichen Confounder gibt. Bei der Datenanalyse ist es notwendig zu wissen, ob das Outcome (z.B. die Myokardinfarktrate) Folge eines Prädiktors (z.B. nicht durchgeführten antihypertensiven Therapie) oder Folge eines Confounders (z.B. Nikotinabusus) ist (Abbildung 4).

## Einleitung



**Abbildung 4: Beispiel für Stratifizierung**

Eine weitere Möglichkeit der Confounderadjustierung ist das Matching. Dabei werden Patienten für eine Studie so ausgewählt, dass es für jeden von ihnen in der jeweils anderen Gruppe, Therapie- oder Kontrollgruppe das Pendant bezüglich der potenziellen Confounder gibt.

Im Allgemeinen ist es allerdings nicht möglich, mit Matching und Stratifizierung für unbegrenzt viele Confounder zu adjustieren, da nicht jeder Patient eine identische Merkmals- bzw. Confounderkombination vorweisen kann. Das heißt, trotz großer Probandenzahlen ist es häufig nicht erreichbar, vergleichbare Partner mit vergleichbarer Merkmalkonstellation zur Gegenüberstellung zu finden. Ebenso können die erzeugten Schichten zum Teil von sehr geringer Größe sein. Sowohl die wenigen Matching-Partner, als auch die vermeintlich kleinen Schichten erhöhen das Risiko des bereits in Kapitel 1.2.2 beschriebenen  $\beta$ -Fehlers mit entsprechender Verringerung der statistischen Trennschärfe <sup>[13]</sup>.

Der methodische Vorteil nicht randomisierter Studien wäre somit verringert bzw. nichtig.

Ein statistisches Modell, dessen Ziel es ist, Beziehungen zwischen Variablen festzustellen bzw. die gleichzeitigen Wirkungen verschiedener Confounder auf ein Outcome zu prüfen, ist die multivariable oder multiple logistische Regressionsanalyse. Es gibt keinen anderen Weg, das Wirken eines Prädiktors auf ein Outcome für mehrere Confounder gleichzeitig zu adjustieren. Die multivariable Modellbildung ist deshalb eine Notwendigkeit in nicht randomisierten, klinischen, Studien.

Als vierte Möglichkeit, die Wirkung eines Prädiktors auf das Outcome für einen Confounder zu adjustieren, sei die Standardisierung zu nennen. Diese findet jedoch eher im Rahmen von Risikostudien ihre Anwendung und soll lediglich der Vollständigkeit halber genannt werden.

### 1.5 Propensity score

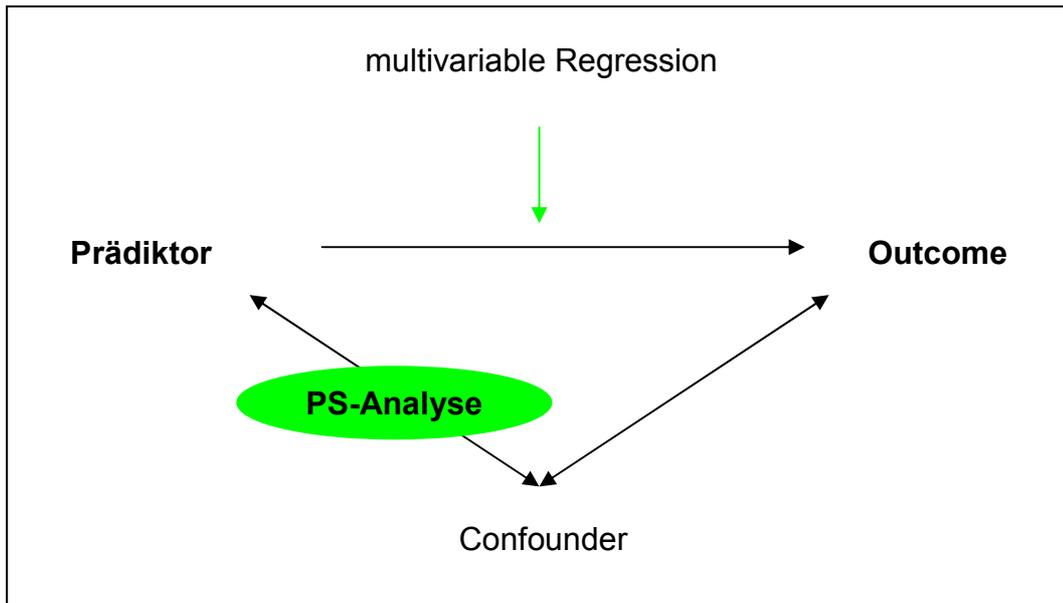
#### 1.5.1 Begriffserklärung

Der 1983 erstmalig von Rosenbaum und Rubin <sup>[22]</sup> vorgestellte Propensity score (PS) ist eine weitere Methode der Confounderadjustierung bei der Datenanalyse nicht randomisierter, kontrollierter, klinischer Studien und hat als wirksame und attraktive Alternative zur multivariablen Regression einige entscheidende Vorteile gegenüber der heutigen Standardmethode <sup>[23; 24; 25]</sup>. Der Anteil angewandter Propensity-score-Analysen stieg im Verlauf des letzten Jahrzehnts stetig an <sup>[5; 26; 27; 28]</sup>. Vor allem in Untersuchungen mit kardiologischen Fragestellungen ist diese Form der Confounderadjustierung zunehmend vertreten <sup>[3]</sup>.

Die Berechnung des Propensity score erfolgt basierend auf den beobachteten Basismerkmalen (Confounder) der Studienteilnehmer mittels logistischer Regression. Der errechnete PS ist ein zusammengefasster Wert für alle bekannten und relevanten Patientenmerkmale, welche einen entscheidenden Einfluss auf die Studienresultate nehmen können. Er stellt per definitionem die Wahrscheinlichkeit dar, als Studienproband in die Therapiegruppe zu gelangen <sup>[5; 29; 30; 31]</sup> und nimmt dementsprechend einen Wert zwischen 0 und 1 an. Es ist anzumerken, dass die Probanden bereits auf die Gruppen verteilt sind.

Im Falle einer Randomisierung bestünde bei jedem Patienten eine fünfzig prozentige Wahrscheinlichkeit, die zu untersuchende Behandlung zu erhalten und deshalb ein PS von 0,5. Im Rahmen von RCTs werden also Patienten mit jeweils einem festen PS von 0,5 verglichen.

Während die multivariable Regression die Prädiktor-Outcome-Beziehung für bekannte Confounder adjustiert, repräsentiert die Propensity-score-Analyse einen Zusammenhang zwischen Confounder und Prädiktor (Abbildung 5).



**Abbildung 5: Unterschied der Angriffspunkte von multivariabler Regression und der Propensity-score Analyse bei der Confounderadjustierung**

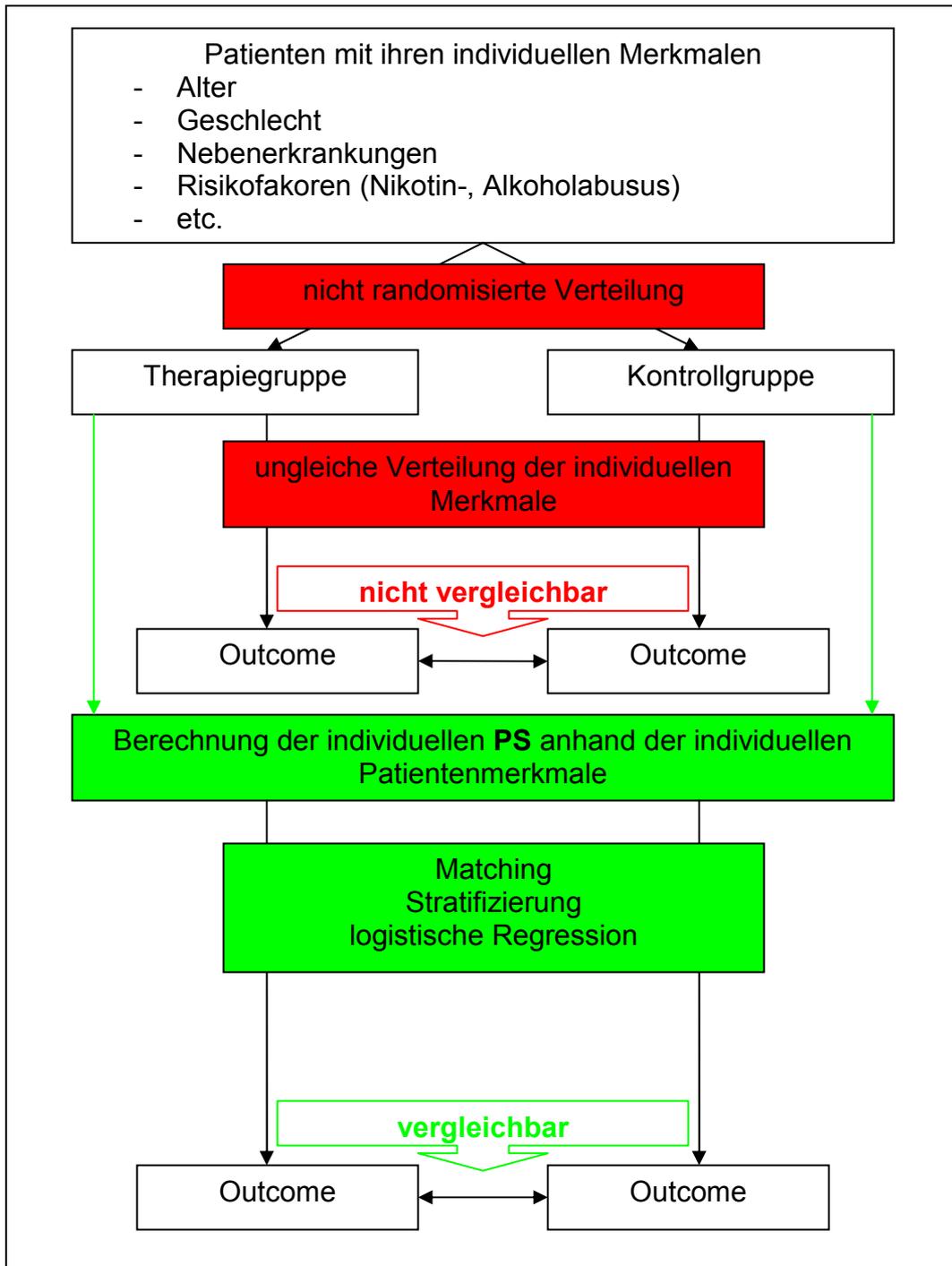
### 1.5.2 Anwendung

Es gibt drei auf dem Propensity score basierende, aber nicht dafür spezifische analytische Techniken, die zur Anwendung kommen <sup>[5]</sup>. Diese finden wie bereits beschrieben, bei der Kontrolle von Selektionsbias bzw. der Confounderadjustierung ihre Anwendung <sup>[13]</sup>.

Beim Matching werden Paare aus Probanden entsprechend ihres individuellen Propensity score gebildet. Man vergleicht zum Beispiel einen Patienten mit einem PS von 0,75 aus der Therapiegruppe mit einem Probanden der Kontrollgruppe, dessen PS ebenfalls 0,75 beträgt.

Der Propensity score kann auch unter Anwendung der Stratifizierung genutzt werden. Bei einer Einteilung in Quintile umfasst die erste Schicht alle Studienteilnehmer mit einem PS zwischen 0 und 0,2, die Zweite von 0,21 bis 0,4, usw. Es werden folglich Probanden aus Therapie- und Kontrollgruppe verglichen, welche sich entsprechend ihrem PS in der gleichen Schicht befinden.

Das dritte Einsatzverfahren des Propensity scores besteht darin, ihn als einzige Variable bzw. einzigen Confounder im Rahmen einer logistischen Regressionanalyse zu nutzen. Damit muss der Einfluss der Intervention auf das Outcome nicht für eine Vielzahl von Variablen, sondern lediglich einen Confounder adjustiert werden.



**Abbildung 6: Schematische Darstellung der Confounderadjustierung unter Verwendung des PS**

Abbildung 6 verdeutlicht den Zusammenhang und die Notwendigkeit der PS-Analyse. Im Rahmen einer nicht randomisierten Beobachtungsstudie obliegt die Verteilung der Studienteilnehmer dem Studienleiter bzw. dem betreuenden Arzt. Diese Problematik der Patientenselektion und damit die wahrscheinlich ungleiche Verteilung der Patientenmerkmale auf die zu vergleichenden Gruppen wurde bereits in den vorangegangenen Kapiteln dieser Arbeit erörtert. Die Abbildung zeigt, dass die PS-Bestimmung erst nach Erhalt der Ergebnisse erfolgt. Nun ist

ein Vergleich, beispielsweise durch Paar- oder Schichtenbildung anhand der Propensity scores möglich und valide.

### 1.5.3 Methodische Vorteile

Angenommen man vergleicht eine konventionelle mit einer neueren, minimalinvasiven Operationsmethode. Die Durchführung einer RCT ist aus ethischen, praktikablen oder finanziellen Anlässen nicht möglich und die Patienten in dem minimalinvasiven Kollektiv sind durchschnittlich jünger und gesünder (beispielsweise aufgrund eines geringeren BMI und kleineren Anteils an Nebenerkrankungen im Sinne von arterieller Hypertonie, Diabetes mellitus und koronarer Herzkrankheit) als in dem konventionellen Kollektiv. Eine weitere Annahme sei, dass nach Studienabschluss und Zusammenschau der erhalten Datensätze ein neues Interesse für ein zusätzliches Outcome, z. B. der Reinterventionsanteil, die Liegezeit oder Kosten, aufkommt. Die entscheidenden Vorteile der Anwendung des Propensity score sind Folgende <sup>[32]</sup>.

Jedem Probanden, der Behandlungs- oder Kontrollgruppe zugehörig, wird, basierend auf seinen individuellen Merkmalen (Confounder), ein PS zugeordnet. Anschließend werden, z. B. im Falle von Matching, nur die Patienten mit identischem PS miteinander verglichen, da nur diese Patienten ähnliche bzw. gleiche Charakteristika besitzen. Es entsteht eine sogenannte „quasi randomisierte“ Situation <sup>[26]</sup>. Ferner ist es möglich, einen einmalig errechneten Wert, auf unzählige separate und sogar nicht zeitlich an den Studienzeitraum gebundene Outcomes anzuwenden, da der PS lediglich die Confounder-Prädiktor-Beziehung repräsentiert.

Ebenso festzuhalten ist die Tatsache, je spärlicher ein Outcome, z. B. die Mortalitätsrate ausfällt, insbesondere in Relation zu den Patientenmerkmalen, die mit in das Modell einfließen, desto fehlerhafter ist die Schätzung durch eine multivariable Modellbildung. Diese sogenannte „Event-per-baseline-characteristic-ratio“ nimmt bei der konventionellen Regressionanalyse einen eher hohen Wert an <sup>[33]</sup>. Es existiert ein Grenzwert von zehn Ereignissen (Events) pro Charakteristikum (Confounder) <sup>[34]</sup>. In einer Studie von Cepeda et al. <sup>[23]</sup>, deren Inhalt die Gegenüberstellung von Propensity-score- und multivariabler Regressionanalyse in nicht randomisierten klinischen Studien in Bezug auf eine geringe Event-Anzahl war, konnte gezeigt werden, dass die PS-Analyse im Sinne

der Confounderkontrolle bessere Ergebnisse erzielte, speziell bei sieben und sogar weniger Events pro „baseline characteristic“. [23]

Die Propensity-score-Analyse hat also einen entscheidenden Vorteil wenn die Konstellation eines seltenen Ergebnisses mit einer Vielzahl an möglichen Einflussfaktoren vorliegt.

### **1.6 Gegenstand der vorliegenden Arbeit**

Da es sich bei der Propensity-score-Analyse um eine sehr valide Methode der Confounderadjustierung handelt und wie oben bereits erwähnt durch diese Methode eine „quasi-randomisierte“ Situation in nicht randomisierten Studien geschaffen wird, liegt ein Vergleich von RCT und PS-Analyse nahe. Wenn man folglich in der Lage wäre, gleiche Ergebnisse mit PS-Analysen zu produzieren und aus diesem Grund die möglichen Differenzen zwischen den Outcomes beider Studienformen aufzuheben, bestünde die Möglichkeit, die Vorteile randomisierter und nicht randomisierter Studien unter Vermeidung der o. g. Nachteile miteinander zu kombinieren. Das Resultat wäre eine hohe interne und externe Validität. Dies würde zum Einen durch eine effektivere Confounderkontrolle verglichen mit der multivariablen Regressionsanalyse realisiert. Zum Anderen trägt eine größere Studienpopulation, eine höhere Trennschärfe, eine längere Verlaufsbeobachtung und eine breitere Vielfalt an Basismerkmalen teilnehmender Probanden im Vergleich zu randomisierten Studien zu einer hohen externen Validität bei. Ein weiterer Vorteil wären die erheblich geringeren Kosten gegenüber der Durchführung einer RCT.

Im Rahmen der vorliegenden Arbeit werden die Ergebnisse randomisierter Studien und Propensity-score-Analysen miteinander verglichen. Der gemeinsame Gegenstand der beiden Studienformen ist der Vergleich konventioneller On-Pump- mit Off-Pump-Operationen in der koronaren Bypasschirurgie.

#### 1.6.1 Problematik des Vergleichs

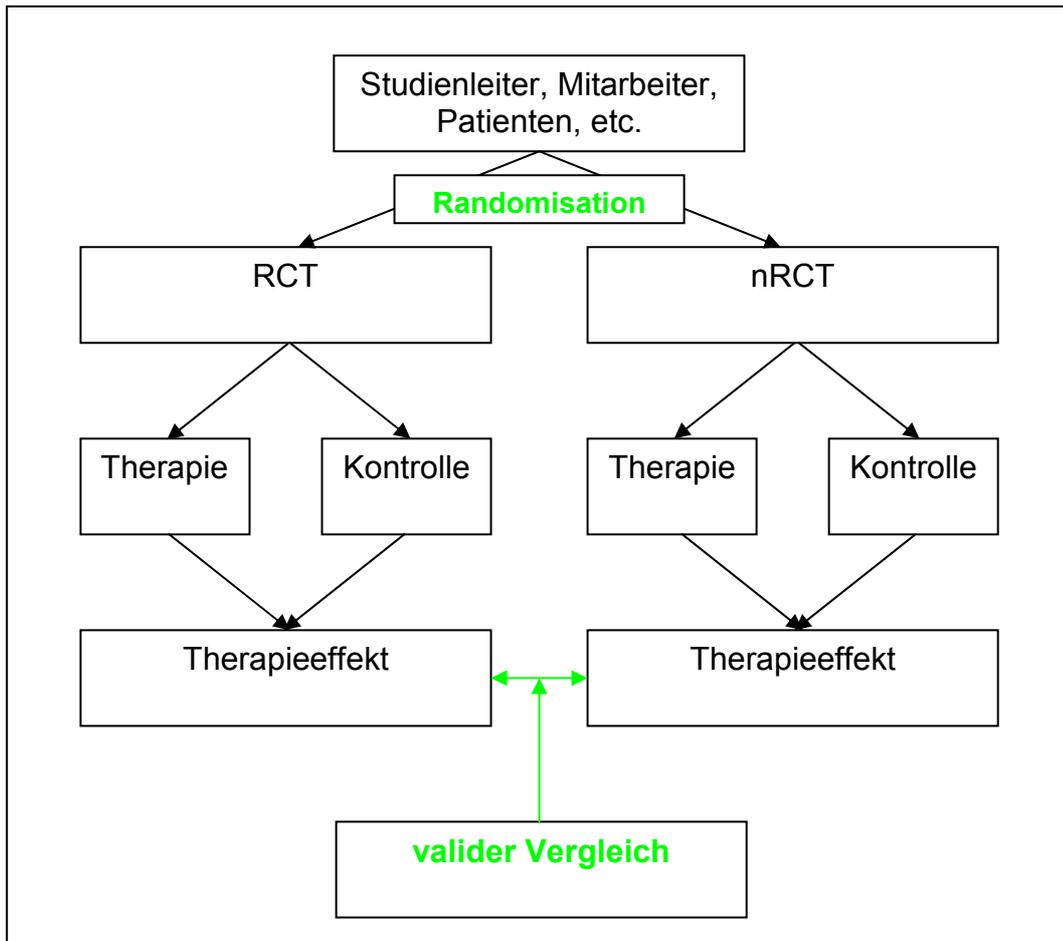
Es sind bisher eine Vielzahl von Untersuchungen durchgeführt worden, die klären sollten, ob sich die aus randomisierten Studien hervorgegangenen Therapieeffekte von denen aus nicht randomisierten Studien unterscheiden. Ein großer Anteil dieser Gegenüberstellungen ist in systematischen Übersichtsarbeiten zusammengefasst worden [12; 36; 37; 38; 39]. Britton et al. [12] vermuteten, dass die Unterschiede zwischen den Studienformen minimal sind,

wenn sichergestellt ist, dass die jeweiligen Studienpopulationen äquivalent und deswegen vergleichbar sind. Deeks et al. <sup>[36]</sup> nannten explizit Meta-Confounder, welche sich typischerweise verzerrend auf Vergleiche von randomisierten und nicht randomisierten Studien auswirken.

Die Problematik des Confoundings existiert demnach nicht nur bei der Gegenüberstellung zweier Populationen, sondern auch bei dem Ergebnisvergleich zweier Studienarten, denn die partizipierenden Probanden, Kliniken und operierenden Kardiochirurgen weisen mit hoher Wahrscheinlichkeit Unterschiede in ihren Eigenschaften bzw. ihren Standards und ihrem Können auf. Es ist anzunehmen, dass, wie oben beschrieben, die Studienteilnehmer einer RCT jünger und gesünder sind als diejenigen einer nRCT. Daneben gibt es Abweichungen bei der bisher durchgeführten Anzahl und demnach der Beherrschung beider Operationsmethoden zwischen den Studienzentren.

Die Erfüllung folgender Punkte wäre für ein Höchstmaß an interner Validität eines derartigen Vergleiches notwendig. Zum Ersten müssen für RCTs und nRCTs gleiche Ein- und Ausschlusskriterien sowohl für die Patienten als auch für die Chirurgen gelten. Bei den Patienten stehen dabei die Basismerkmale, z. B. Alter, Geschlecht, Komorbidität oder kardiovaskuläre Risikofaktoren im Mittelpunkt. Bei den Chirurgen müssen die bisherige Anzahl an durchgeführten Bypassoperationen, die Technik an sich sowie das verwendete Material, beispielsweise die Stabilisatoren bei Off-Pump-Operationen, identisch sein. Zum Zweiten ist es notwendig, die Probanden und behandelnden Ärzte sowie deren Basismerkmale entsprechend dem operativen Können gleich auf die Studienformen zu verteilen und somit die Beziehung zwischen der Operationsmethode (Prädiktor) und dem Outcome für mögliche Confounder zu adjustieren.

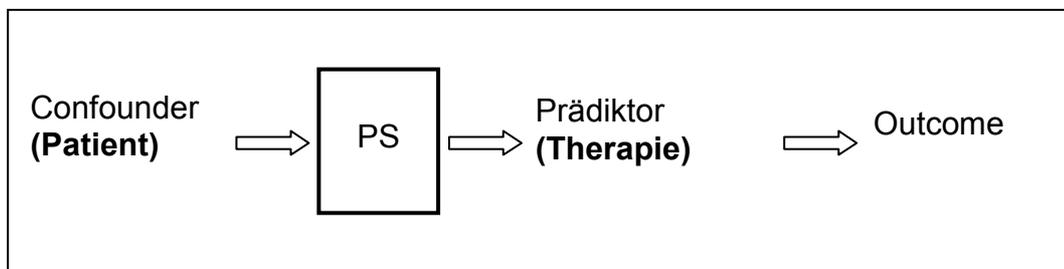
Idealerweise würde man für einen systematischen Vergleich eine „meta-randomisierte“ Studie durchführen (Abbildung 7).



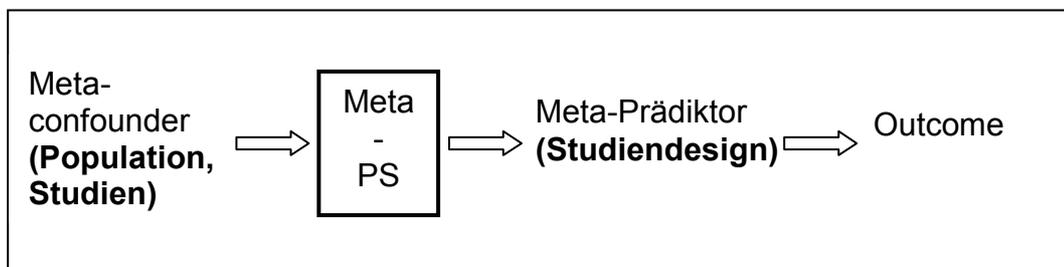
**Abbildung 7: Meta-Randomisation**

Die Studienleiter und ihre Mitarbeiter, welche eine Studie zu einer bestimmten klinischen Fragestellung durchführen wollen, würden einer Randomisierung unterzogen werden, um festzulegen, welche Studienform (RCT oder nRCT) sie durchführen sollen. Dabei verteilt man die potenziellen, sowohl bekannten als auch nicht bekannten Meta-Confunder gleichmäßig auf die Gruppe der RCTs und nRCTs. Alle möglichen Meta-Confunder würden somit eliminiert bzw. kontrolliert werden. Die Meta-Randomisation ist jedoch aus technischen und finanziellen Gesichtspunkten nur mit erheblichem Aufwand realisierbar. Ob eine Realisierung überhaupt möglich ist, hängt auch stark von der Fragestellung ab. Vor allem im medizinischen Bereich wäre eine solche Untersuchung nicht in jedem Fall ethisch vertretbar. Es wäre unethisch, einen Studienleiter, welcher sich für die Umsetzung einer randomisierten Studie entschieden hat mittels Meta-Randomisation zu zwingen, eine nicht randomisierte Studie durchzuführen. Insbesondere bei der Erforschung neuerer, vor allem lebensverlängernder Therapiemethoden ist ein Studienleiter angehalten, den aktuellen Qualitätsstandard, mit einem Höchstmaß an interner Validität zur Bewertung dieser Therapiemethoden zu benutzen.

Da davon auszugehen ist, dass der Propensity score die valideste Analysemethode ist, nicht randomisierte Studien auszuwerten, wird im Rahmen dieser Arbeit eine Meta-Propensity-score-Analyse praktiziert, um eventuelle Ergebnisunterschiede zwischen RCT und nRCT eruieren und evaluieren zu können. Die Confounder auf Studienebene werden als „Meta-Confounder“ bezeichnet. Es wird eine Beziehung konstruiert, die Auskunft darüber gibt, mit welcher Wahrscheinlichkeit eine Studie aufgrund ihrer eingeschlossenen Patienten und Studieneigenschaften (Multizenterstudie, Verblindung, etc.) eine RCT oder PS-Analyse sein kann. Somit erhält jede einzelne Studie anhand der Grundmerkmale auf Studien- und Probandenebene einen individuellen Meta-PS. Anschließend erfolgt ein Meta-Matching. Es werden die RCTs und nRCTs miteinander verglichen, die einen identischen Meta-PS besitzen (Abbildung 8, 9).



**Abbildung 8: Prinzip der Propensity-Score-Analyse**



**Abbildung 9: Prinzip der Meta-Propensity-Score-Analyse; vgl. Abb. 8**

### 1.7 Aortokoronare Bypassoperation

#### 1.7.1 Zugangsweg <sup>[40]</sup>

Die mediane Sternotomie dient als Standardzugang für diesen kardiochirurgischen Eingriff. Die Schnittlänge kann variiert werden. Auch eine Reduktion der Inzisionen auf 10 – 15 cm führt nicht zu einer wesentlichen Sicherheitseinschränkung. Die mediane Sternotomie ist der einzige Zugangsweg, über den die Graftgewinnung der Aa. thoracicae internae bis zu deren Abgängen aus der entsprechenden A. subclavia erfolgen kann. Eine Zugangsalternative ist die anterolaterale Minithorakotomie im Bereich des linken 4.-5. Interkostalraumes. Diese MIDCAB-Technik (Minimally invasive direct coronary artery bypass) fand in den für diese Arbeit genutzten Studien keine Anwendung.

#### 1.7.2 On-Pump-Technik <sup>[41]</sup>

In Deutschland werden über 85% der Bypassoperationen in On-Pump-Technik durchgeführt <sup>[42]</sup>. Der Patient wird dabei in Rückenlage gebracht. Die Arme werden anliegend oder in 90° Abduktion gelagert. Der Operateur steht rechtsseitig, der erste Assistent linksseitig vom Patienten. Nach erfolgter Intubationsnarkose durch den Anästhesisten, Lagerung des Patienten sowie sterilem Abwaschen und Abdecken erfolgt die mediane Sternotomie.

Bei der Standardmethode in der aorto-koronaren Bypasschirurgie werden die A. thoracica interna sinistra (LIMA: left internal mammarian artery) sowie Veneninterponate (V. saphena magna) verwendet. Die Präparation der LIMA kann auf zwei verschiedenen Wegen erfolgen. Zum einen wird ein Pedikel präpariert. Dieser beinhaltet LIMA, Pleura parietalis, Muskelfascie und begleitende Venen. Die Zweite Möglichkeit ist die separate Freilegung der LIMA. Nach Vollheparinisierung des Patienten (400IE/kg KG) wird die LIMA distal durchtrennt, mit einem Clip verschlossen und zur Aufbereitung in eine mit Heparin gefüllte 20ml Spritze eingetaucht. Eine perivasale Infiltration mit einer 1:10 verdünnten Papaverin-Heparin-Ringer-Lactatlösung kann mögliche Gefäßspasmen verhindern. Eine direkte intraluminale Injektion sollte wegen einer möglichen Intimaschädigung unterlassen werden. Parallel zur LIMA-Präparation wird am Bein des Patienten das Veneninterponat gewonnen. Die gewonnene V. saphena magna wird mit isotoner Kochsalzlösung mit maximalen Druck von 150mmHg gespült und aufgeweitet. Die dabei verwendete Knopfkanüle wird an

das distale Ende der Vene eingebunden, um die Orientierung zur Flussrichtung für den späteren Bypass sicherzustellen.

Der konventionelle koronar-arterielle Bypass (CCAB) am stillgelegten Herzen wird unter Hypothermie (32 bis 35°C) durchgeführt. Durch die Hypothermie kann eine Drosselung des Stoffwechsels und somit eine Erhöhung der Ischämietoleranz aller Organe erreicht werden. Dafür ist die Verwendung einer Herz-Lungen-Maschine zur extrakorporalen Zirkulation notwendig. Für den Anschluss der HLM stehen dem Operateur verschiedene Kanülierungstechniken zur Verfügung. Zur Standardmethode gehören die Kanülierung der Aorta ascendence (distal der Koronararterien) mit dem Kardioplegiekatheter sowie des proximalen Aortenbogens für den Rückfluss des oxygenierten Blutes aus der HLM. Die Kanülierung des venösen Schenkels erfolgt entweder separat für V. cava superior und inferior oder mit einem Zweistufenrückflusskatheters an der Basis des rechten Herzhohls. Anschließend wird durch Infusion einer auf 4°C abgekühlten kristalloiden Kardioplegielösung<sup>13</sup> proximal der Aortenabklemmung und distal der Koronarostien das Herz zum Stillstand gebracht. Durch die Temperatur der Kardioplegielösung wird das Herz auf circa 10°C abgekühlt. Damit wird die Ischämietoleranz des Myokards verlängert.

Die Präparation der Koronargefäße muss mit großer Sorgfalt erfolgen. Es sollte vor dem kardioplegischen Arrest des Herzens eine ausführliche Inspektion der Gefäße erfolgen, um das Verhalten der Gefäße beim Herzschlag einschätzen zu können.

In einigen Fällen zeigen die Koronargefäße einen intramyokardialen Verlauf. In diesen Fällen ist es zwingend erforderlich, während des myokardialen Arrestes diese Gefäße sorgfältig freizupräparieren. Das Myokard im diesem Bereich muss mit äußerster Vorsicht inzidiert werden bis die Gefäßwand erreicht ist. Intramyokardial verlaufende Gefäße haben häufig eine viel dünnere Wand als die epikardial verlaufenden. Die entsprechende Koronararterie wird distal der Stenosierung inzidiert. Die Inzision wird mit einer Koronarschere verlängert. Das vorher gewonnen Transplantat wird schräg angeschnitten und anschließend mit einem 7/0 Polypropylenfaden fortlaufend fixiert. Die Anastomisierung kann End-zu-Seit oder Seit-zu-Seit bei der Anlage von sequentiellen Bypassen (jump

---

<sup>13</sup> Es werden Kardioplegielösungen von Buckberg, St. Thomas oder Calafiore verwendet. Kalium und Magnesium sorgen durch schnelle Depolarisation über der Kardiomyozytenmembran zu einem diastolischen Herzstillstand. Die Lösungen unterscheiden sich gering in ihren Zusammensetzungen. Sie werden mit einer Lösungstemperatur von 4°C -8°C bzw. Calafiore mit 37°C appliziert <sup>[35; 43; 44]</sup>.

grafts) durchgeführt werden. Die Anastomisierung der LIMA, in aller Regel mit dem Ramus interventricularis anterior (RIVA oder LAD: left anterior descending) erfolgt End-zu-Seit. Anschließend wird die Perfusion durch Lösung der Aortenklemme freigegeben, was in der Regel zu einer spontanen Wiederaufnahme der Herztätigkeit führt. Bei eventuellem Herzflimmern besteht die Notwendigkeit einer Defibrillation. Um den endgültigen Aortocoronaren Venenbypass (ACVB) zu erhalten, wird mit einer Seitenklemme ein Teil der Aorta längs ausgeklemmt. Mit einer Stanze wird die Aortenwand kreisrund eröffnet. Mit einer fortlaufenden Naht (7/0 Polypropylenfaden) werden die Veneninterponate mit der Aorta verbunden und die ausgeklemmte Aorta wieder freigegeben. Es erfolgt eine Entlüftung der Grafts mittels kleiner Injektionsnadeln und das Myokard wird wieder perfundiert. Nach einer Reperfusionphase wird die extrakorporale Zirkulation durch die HLM langsam beendet. Danach wird Protamin zur Heparinneutralisation im Verhältnis Protamin : Heparin 1 : 1 appliziert. Am rechten Vorhof und rechten Ventrikel anliegende Schrittmacherelektroden können zur eventuellen Stimulation eingesetzt werden. Es erfolgt ein schichtweiser Verschluss des Operationsgebietes.

### 1.7.3 Off-Pump-Technik <sup>[45; 46]</sup>

2006 wurden in Deutschland 5294 Off-Pump-Bypassoperationen durchgeführt <sup>[42]</sup>. Das ist eine Anteilzunahme um 2,4% aller isolierten koronarchirurgischen Eingriffe im Vergleich zum Jahr 2004 <sup>[47]</sup>. Das Perioperative anästhesiologische Management nimmt bei der Durchführung der Off-Pump-Technik einen sehr hohen Stellenwert ein, um eine hämodynamische Stabilität des Patienten und damit den Erfolg der OPCAB-Operation zu gewährleisten. Zum differenzierten Einsatz kommen dabei Vasodilanzien, Volumengabe, Inotropica und vorübergehendes Vorhofpacing. Aus chirurgischer Sicht gleicht die Bypassoperation am schlagenden Herzen in Bezug auf Zugangsweg, verwendeten Bypassmaterialien (LIMA, Veneninterponat) und deren Gewinnung der CCAB-Technik.

Es sind verschiedene Stabilisierungstechniken notwendig, um eine identische Präzision der Anastomosennaht am schlagenden Herzen erreichen zu können. Die Darstellung der Koronargefäße wird durch Luxation des Herzens aus dem Perikardbeutel mittels tiefer Perikardzugnähte und variabler Retraktionsbänder ermöglicht. Für die Exposition posterolateraler Areale kommen Stabilisierungssysteme zum Einsatz. Diese umfassen ähnlich einer Saugglocke

die Herzspitze, um diese atraumatisch anheben zu können. Die Vorlast des Herzens muss dabei durch Infusion von 1,5-2l Flüssigkeit und Kopftieflage des Patienten durch den Internisten erhöht werden. Ein Stabilisator mit angeschlossenem Unterdrucksystem schafft ein ruhiges und sicheres Operationsfeld im Bereich der Anastomosennaht. Nach Heparinisierung mit einem Drittel der Vollheparinisierungsdosis wird wenige Millimeter distal der geplanten Koronararterieninzision eine Umstechungsnah angelegt. Mit Hilfe eines Tourniquets kann nun die Koronardurchblutung gedrosselt werden, um ein blutfreies Sichtfeld zu schaffen. Calciumantagonisten und  $\beta$ -Blocker können dabei zu Erhöhung der Ischämietoleranz genutzt werden. Jetzt kann die Koronararterie inzidiert werden. Durch die Anwendung eines CO<sub>2</sub>-Mikroblowers kann zusätzlich noch die Sicht einschränkendes Blut vorsichtig weggeblasen werden. Die Inzision wird mit einer Koronarschere erweitert. Eine weitere Möglichkeit ein blutfreies Anastomosengebiet zu erzielen sind intrakoronare Shunts. In die eröffnete Koronararterie wird ein circa 1-2mm dickes Polyurethanröhrchen eingeschoben. Dabei wird der Blutaustritt vermieden, jedoch die Koronardurchblutung weitestgehend erhalten. Analog zur CCAB-Technik erfolgt mit 7/0 Polypropylenfaden fortlaufend die Anastomosennaht. Vor dem Anziehen der letzten Nahtschlinge wird der Shunt entfernt. Durch das Anbringen der Aortenklemme wird entsprechend der oben beschriebenen On-Pump-Technik der ACVB fertiggestellt. Anschließend erfolgt die Antagonisierung des Heparins mittels Protaminsulfat. Das Operationsgebiet wird schichtweise verschlossen.

## **2. Zielsetzung der vorliegenden Arbeit**

### **2.1 Fragestellung**

Das Ziel der vorliegenden Arbeit ist es, die Ergebnisse randomisierter, kontrollierter, klinischer Studien mit denen nicht randomisierter, klinischer Beobachtungsstudien, in denen der Propensity score als Analyseverfahren zur Anwendung gekommen ist, miteinander zu vergleichen. Dieser Vergleich soll zeigen, ob es einen bzw. keinen signifikanten Unterschied in den Studienergebnissen gibt, wenn nach erfolgtem Meta-Matching beide Studienformen ähnliche Populationen haben. Es soll gezeigt werden, ob die Nutzung der Propensity-score-Analyse in Anbetracht der Ergebnisvalidität dem „Gold-Standard“ der epidemiologischen Modelle ebenbürtig oder sogar überlegen ist.

### **2.2 Hypothesen**

1. Es gibt keinen Unterschied in den gemessenen Therapieeffekten zwischen randomisierten, kontrollierten, klinischen Studien und nicht randomisierten Studien, in denen der Propensity score als Analyseverfahren Anwendung fand.
2. Nach erfolgtem Meta-Matching ist bezüglich der Meta-Confounder kein signifikanter Unterschied zwischen den Studienformen erkennbar.
3. Ergebnisunterschiede zwischen RCT und nRCT sind nicht auf die fehlende Randomisation, sondern auf die unterschiedlichen Studienpopulationen zurückzuführen.
4. Das Patientenkontext der randomisierten Studien ist durchschnittlich jünger und gesünder als das der nicht randomisierten Studien.
5. Sowohl randomisierte Studien als auch PS-Analysen zeigen einen signifikanten, positiven Therapieeffekt auf der Seite der mittels Off-Pump-Technik durchgeführten Bypassoperationen.

## **3. Material und Methodik**

### **3.1 Studienthemen**

Der gemeinsame Gegenstand der zu vergleichenden randomisierten und nicht randomisierten, klinischen Studien ist die Gegenüberstellung von Bypassoperationen am kardioplegischen Herzen unter Verwendung einer Herz-Lungen-Maschine (On-Pump; CABG) und Bypassoperationen am schlagenden Herzen ohne Anschluss eines extrakorporalen Umgehungskreislaufes (Off-Pump; OPCAB). Dieses Thema ist das derzeit wohl am meisten debattierte und am meisten polarisierende im Bereich der Herzchirurgie <sup>[48]</sup>. In Deutschland wurden im Jahr 2006 64.502 Bypassoperationen durchgeführt. In 8,2% der Fälle operierten die Chirurgen unter Verwendung der Off-Pump-Technik <sup>[42]</sup>.

### **3.2 Studiensuche und Datenextraktion aus RCTs**

#### **3.2.1 Studiensuche**

Die Literaturverzeichnisse der fünf größten systematischen Übersichtsarbeiten zum Thema Off-Pump versus On-Pump von Cheng et al. <sup>[49]</sup>, Raja <sup>[50]</sup>, Reston et al. <sup>[51]</sup>, Sedrakyan et al. <sup>[52]</sup> und Wijesundera et al. <sup>[53]</sup> wurden als Quellen genutzt, um die randomisierten, kontrollierten, klinischen Studien zusammenzutragen. Die aktuellste Übersichtsarbeit von Sedrakyan schloss jedoch lediglich bis einschließlich Februar 2006 publizierte Studien ein. Aus diesem Grund wurde im Oktober 2006 zusätzlich eine MEDLINE-Suche mit den Schlagwörtern „randomized“ und „off-pump“ durchgeführt.

#### **3.2.2 Datenextraktion**

Nach Erhalt der Volltexte wurden diese unabhängig voneinander (Dr. rer. nat. Oliver Kuss, Thomas Legler) und unter Verblindung für die Ergebnisse der fünf Metaanalysen gelesen. Mit Hilfe eines fünfseitigen Auswertungsbogens (Anhang 8.1) wurden die folgenden Daten extrahiert.

- I. Generelle Informationen zur vorliegende Studie

- a. Titel, Autor, Quelle

- II. Einschluss- bzw. Ausschlusskriterien der vorliegenden Studie

- a. Findet ein Vergleich zwischen einer Off-Pump- und einer On-Pump-Gruppe statt?
- b. Handelt es sich bei der vorliegenden Studie um eine RCT?
- c. Ist der Therapieeffekt bezüglich eines der 10 binären Outcomes (s.u.) in der vorliegenden Studie untersucht worden?
- d. Handelt es sich um primär und nicht aus systematischen Übersichten oder Metaanalysen gewonnene Daten?

### III. Kategorische Meta-Confounder

- a. Beobachtungszeitraum
- b. Studienregion
  - i. Europa
  - ii. Nordamerika
  - iii. Andere Region
- c. Anzahl der Studienzentren
  - i. Unizenterstudie
  - ii. Multizenterstudie
  - iii. Fehlende Angabe
- d. Anzahl randomisierter und analysierter Patienten
- e. Cross-over:  
Patientenanzahl, die von der Off- in die On-Pump-Gruppe wechselten (in % als Anteil der Gesamtstudienpopulation)

### IV. Kontinuierliche / Stetige Meta-Confounder

- Patientenmerkmale und Risikofaktoren

### V. Dichotome Ergebnisse

- a. Absolute Anzahl der Events
  - i. Mortalität postoperativ
  - ii. Apoplex postoperativ
  - iii. Myokardinfarkt postoperativ
  - iv. Vorhofflimmern postoperativ

- v. Transfusionen von Erythrozytenkonzentraten
- vi. Katecholaminpflicht
- vii. Wundinfektionen
- viii. Revisionsoperation infolge einer Nachblutung
- ix. akute Niereninsuffizienz postoperativ
- x. Einsatz von IABP

Nach Beendigung der unabhängigen Datensammlung wurde diese in eine MS ACCESS Datenbank eingetragen und automatisch auf Unstimmigkeiten überprüft. Die anschließende Korrektur erfolgte entweder durch einen einfachen Konsens, durch eine Diskussion mit einem dritten Mitarbeiter, einem Kardiochirurgen, oder, als Ultima Ratio, durch Zuhilfenahme der zuvor genannten systematischen Übersichtsarbeiten von Cheng, Raja, Reston, Sedrakyan oder Wijeyesundera.

### 3.2.3 Studieneinschluss

Ein Einschluss einer RCT in diese Arbeit erfolgte nur bei eindeutig positiver Beantwortung aller vier zuvor unter II. angegebenen Fragen. Sichere und vermeintliche Doppelpublikationen wurden ausgeschlossen, jedoch die zum Teil zusätzlich zur bereits eingeschlossenen Studie enthaltenen Informationen zu den Probanden und den Ergebnissen wurden mit in den für den Vergleich relevanten Datensatz aufgenommen.

## 3.3 Datensuche und Datenextraktion aus PS-Analysen

### 3.3.1 Studiensuche

Es wurden PS-Analysen aus einer Meta-Analyse mehrerer bis Februar 2006 erschienener, nicht randomisierter, klinischer Beobachtungsstudien, in denen der PS zur Datenauswertung Anwendung fand, in die hier beschriebene Meta-Propensity-score-Analyse eingeschlossen <sup>[54]</sup>.

Die Studiensuche wurde basierend auf drei Pfeilern von zwei Personen unabhängig von einander in der fünften Kalenderwoche des Jahres 2006 durchgeführt.

Zum Ersten wurden mit Hilfe der Schlüsselwörter „Propensity“ und „Off-Pump“ die Literaturdatenbanken von MEDLINE, EMBASE, ACP Journal Club, CCTR (Cochrane Central Register of Controlled Trials), CDSR (Cochrane Database of Systematic Reviews), DARE (Database of Abstracts of Reviews of Effects), EBM Reviews und Web of Science ohne genauere Angabe bzw. Einschränkung der Publikationszeiträume durchsucht.

Des Weiteren wurde in den Zitaten und Literaturverzeichnissen von sechs Publikationen <sup>[5; 22; 26; 31; 55; 56]</sup> über die Methodik des Propensity scores via Web of science nach weiteren Studien gesucht.

Zum Dritten fand eine Suche mittels selbigen oben genannten Schlüsselwörtern in den wissenschaftlichen Datenbanken von Google Scholar (<http://scholar.google.com>), Scirus (<http://www.scirus.com>) und Vivisimo clustering (<http://vivisimo.com>) statt.

Zusätzlich wurden alle Literaturverzeichnisse der erhaltenen Publikationen nach weiteren Studien überprüft, welche die konventionelle und die Off-Pump-Technik bei koronaren Bypassoperationen gegenüberstellten. Neben vollständigen Artikeln wurden auch Zusammenfassungen und nicht veröffentlichte Berichte beachtet. Die Autoren wurden angeschrieben, um fehlende Informationen ihrer Ausführungen zu liefern.

Im Oktober 2006 wurde eine erneute Suche identischen Schemas nach jüngst publizierten PS-Analysen durchgeführt.

### 3.3.2 Datenextraktion

Die Datenextraktion erfolgte ebenfalls unter Nutzung eines dem bei der Analyse der RCTs zur Anwendung gekommenen, ähnlich konzipierten Auswertungsbogens.

Essentiell für den Studieneinschluss waren Angaben zu den oben beschriebenen kategorischen und stetigen Meta-Confoundern und zu mindestens einem der zehn für die Evaluation notwendigen binären Outcomes.

### 3.3.3 Studieneinschluss

Für den Einschluss einer Propensity-Score-Analyse müssen die folgenden Kritikpunkte erfüllt sein. Es muss eine On-Pump- mit einer Off-Pump-Gruppe

verglichen werden. Aus diesem Vergleich müssen Informationen zu mindestens einem unter 3.2.2 genannten binären Outcomes und Angaben zu den Meta-Confoundern (Patienten- und Studieneigenschaften) hervorgehen. Doppelpublikationen wurden ausgeschlossen. Jedoch wurden zusätzliche Informationen für diese Arbeit extrahiert und genutzt.

### 3.4 Statistische Methoden

#### 3.4.1 Datenbeschreibung

Zur Beschreibung der kategorischen Meta-Confounder werden die absoluten und relativen Häufigkeiten angegeben. Alle stetigen Meta-Confounder werden unter Angabe der Mittelwerte, dem prozentualen Anteil und der absoluten Anzahl der Beobachtungen in der respektiven Gruppe aufgeführt. Die dazugehörigen p-Werte zur statistischen Signifikanzprüfung wurden mittels  $\chi^2$ -Test, bezogen auf die kategorischen und t-Test, bezogen auf die kontinuierlichen Meta-Confounder, berechnet. Die standardisierten Differenzen werden als Ausdrucksform für die Unterschiede in den Patientenmerkmalen angegeben. Wird eine standardisierte Differenz von 10% unterschritten, gilt die Annahme einer balancierten Verteilung der Charakteristika auf beide Studienformen<sup>[57]</sup>.

Für jedes extrahierte, klinische Outcome wurde die absolute Anzahl der beobachteten Events angegeben.

#### 3.4.2 Umgang mit unterschiedlichen Ausdrücken der Verteilungsbeschreibung

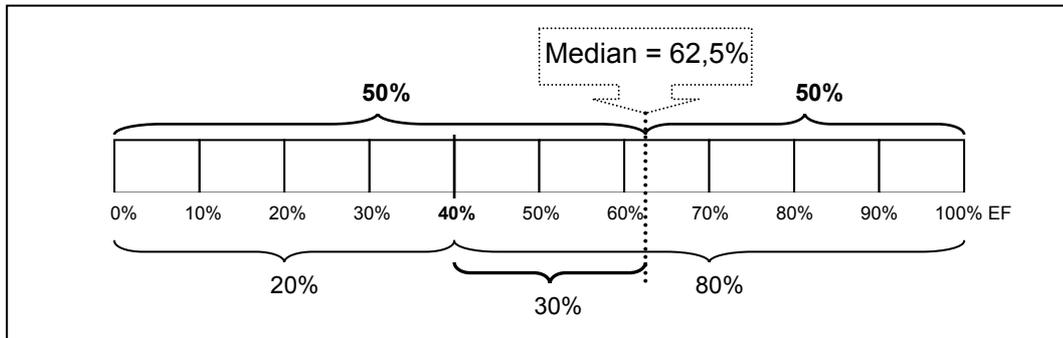
Der Median und der Mittelwert wurden in nahezu identischer Häufigkeit als Verteilungsbeschreibung der stetigen Meta-Confounder benutzt.

In dieser Arbeit wurden zwei Verfahren zum Umgang mit den verschiedenen Kennziffern der zentralen Tendenz gebraucht.

Median und Mittelwert wurden als beschreibende Größen gleichgesetzt, da anzunehmen ist, dass beide zufällig und demnach nahezu identisch auf die RCT und nRCT verteilt sind.

Für den Fall, dass die stetigen Meta-Confounder in verschiedenen Kategorien einer prozentualen Verteilung dargestellt wurden, galt die Annahme einer Gleichverteilung der Werte auf die einzelnen Kategorien und es wurde aus den gegebenen Größen ähnlich dem folgenden Beispiel der Median berechnet.

Eine Studie gibt an, dass bei 80% der Patienten eine LVEF von größer als 40% vorliegt. Der Median der verteilten Ejektionsfraktion würde bei  $[(80\% - 50\%) / 80\% \times 60\% - 40\%] = 62,5\%$  liegen (Abbildung 10).



**Abbildung 10: Graphische Darstellung des Prinzips der Medianberechnung**

### 3.4.3 Umgang mit fehlenden Daten

Es war vor auszusehen, dass nicht jeder potenzielle Meta-Confounder aus allen gelesenen Studien extrahiert werden konnte. Somit wurden nur Patientenmerkmale und prognostische Faktoren als Meta-Confounder in das Meta-PS-Modell eingesetzt, wenn sie in mindestens zwei Drittel aller Publikationen existent waren. Mit den noch fehlenden Werten wurde wie folgt umgegangen.

Die kategorischen Meta-Confounder wurden in gesonderten Gruppen zusammengefasst. Folglich bildeten fehlende Angaben, z.B. unklare Studienregion oder Anzahl der beteiligten Studienzentren, jeweils eine gemeinsame, separate Gruppe.

Für die fehlenden stetigen Meta-Confounder wurde die „Multiple-Imputationsmethode“ benutzt. Es wurden Datensätze erzeugt, in denen kraft „fiktiver“ Regelsätze „fiktive“ Werte erzeugt wurden, welche die nicht gelisteten ersetzen. Dies soll an einem folgenden, stark vereinfachten Beispiel erörtert werden. In vier Studien ist das durchschnittliche Lebensalter der Patienten angegeben, 73 Jahre in Studie 1, 71 Jahre in Studie 2, 70 Jahre in Studie 3 und 68 Lebensjahre in Studie 4. Die durchschnittliche linksventrikuläre Ejektionsfraktion sei in 1 68%, in 2 60%, in 4 58% und in 3 nicht aufgeführt. Man kann dementsprechend zwei Wertereihen gegenüberstellen, 73 – 71 – 70 – 68 für das Probandenalter und 68 – 60 – X – 58 für die kardiale linksventrikuläre Auswurf fraktion. Demgemäß könnte oder müsste X als fehlender stetiger Meta-

Confounder der betreffenden Studie einen Wert zwischen 58 und 60, um sich synergistisch den zugehörigen Lebensaltern zu verhalten. Die erhaltene LVEF wäre ein imputierter Wert und die erhaltenen Datensätze Imputationsdatensätze, wovon Tausende pro Studie, mit jeweils einem eingesetzten Patientenalter für X, generiert wurden, um dem fehlenden Wert möglichst nahe zu kommen. Aus den für jeden dieser Imputationsdatensätze ermittelten Meta-Propensity-Scores wurde der die jeweilige Studie repräsentierende mediane Meta-PS gebildet, der beim nachfolgenden Meta-Matching zum Einsatz kam.

### 3.4.4 Meta-Propensity-score-Modell

Das Meta-PS-Modell wurde mittels logistischer Regression mit einer variablen funktionalen Form der Meta-Confounder berechnet. Funktionale Form bedeutet, dass nach Inspektion der angepassten Kurven in einem semiparametrischen Modell ersichtlich wurde, dass es sich nicht nur um linear, sondern um zum Teil kubisch oder quadratisch verhaltende Beziehungen zwischen Zielgröße und Meta-Confounder handelt. Um die Studienform – Meta-Confounder – Beziehung für den Meta-Propensity-Score zur adjustieren, wurde ein Optimal-Matching-Algorithmus mit einer variablen Anzahl an erlaubten Paaren (Matches) angelegt.<sup>[58]</sup> Für den Matching-Prozess wurden nicht die Meta-Propensity-scores eingesetzt, sondern deren korrespondierende Meta-Prädiktoren.

Die beschriebenen Berechnungen wurden mittels SAS® V9.1 durchgeführt.

### 3.4.5 Klinische Zielgrößen

Für den Vergleich der Behandlungseffekte beider Therapiemethoden und damit der Studienformen wurden Differenzen der logarithmischen Odds Ratios und die 95%-Konfidenzintervalle herangezogen. Diese wurden aus mittels Meta-Confoundern und Effektschätzern konstruierten Vierfeldertafeln ermittelt. Die unterschiedliche Studienwichtung erfolgte mit Hilfe der Inversen-Varianz-Methode. Anschließend wurden die Differenzen der Odds Ratios der jeweiligen klinischen Zielgröße aller Propensity-Score Analysen und randomisierten Studien, sowohl im gematchten als auch im ungematchten Sample gebildet, um mögliche Unterschiede darstellen zu können.

## 4. Ergebnisse

### 4.1 Studieneinschluss

#### 4.1.1 Einschluss der randomisierten, kontrollierten, klinischen Studien

Unter Zuhilfenahme der Übersichtsarbeiten von Cheng et al. <sup>[49]</sup>, Raja <sup>[50]</sup>, Reston et al. <sup>[51]</sup>, Sedrakyan et al. <sup>[51]</sup> und Wijeyesundera et al. <sup>[53]</sup> konnten 81 und mit der zusätzlichen MEDLINE-Suche im Oktober 2006 weitere 26 Publikationen von vermeintlich randomisierten Studien zum Thema CABG versus OPCAB auffindig gemacht werden. Aus diesen 107 gelesenen Artikeln fanden lediglich 51 aus nachfolgenden Gründen in dieser Arbeit Verwendung (vgl. Abbildung 11).

Der Ausschluss von 36 Artikeln erfolgte wegen Doppelpublikation, von zehn wegen fehlender Beschreibung der o. g. zehn binären Outcomes, von acht anlässlich des fehlenden Vergleiches einer On-Pump- mit einer Off-Pump-Gruppe, von weiteren acht, weil sie über die Ergebnisse aus nicht randomisierten Studien berichteten. Bei einer der 107 Publikationen handelte es sich um eine systematische Übersichtsarbeit und bei einer Studie lag keine ausreichende Beschreibung der Meta-Confounder vor.

Von den ausgeschlossenen Doppelpublikation handelte es sich bei zwei der 36 zusätzlich um nRCTs statt der notwendigen RCTs, bei drei Ausführungen um Studien ohne Angabe zu den für diese Arbeit wichtigen Outcomes und bei einer nicht um einen Vergleich von CABG und OPCAB. Bei einer Arbeit fehlte neben einer Beschreibung der Outcomes auch eine Randomisation der teilnehmenden Probanden.

Aus vorangegangenen Gründen und der Festlegung, dass es zulässig ist, zusätzliche Daten über die kategorischen und stetigen Meta-Confounder sowie den Outcomes aus doppelt veröffentlichten Studien zu nutzen, konnten Informationen zu 51 eingeschlossen aus insgesamt 57 Publikationen gewonnen werden.

Nachfolgend soll ein Flussdiagramm (Abbildung 11) über den Prozess der Studienelektion informieren. Außerdem sind zur Übersicht sowohl die eingeschlossenen als auch die ausgeschlossenen Studien tabellarisch mit Erstautor, Herkunft und Ausschlussgrund im Anhang aufgelistet (Anhang 8.2 und 8.3).

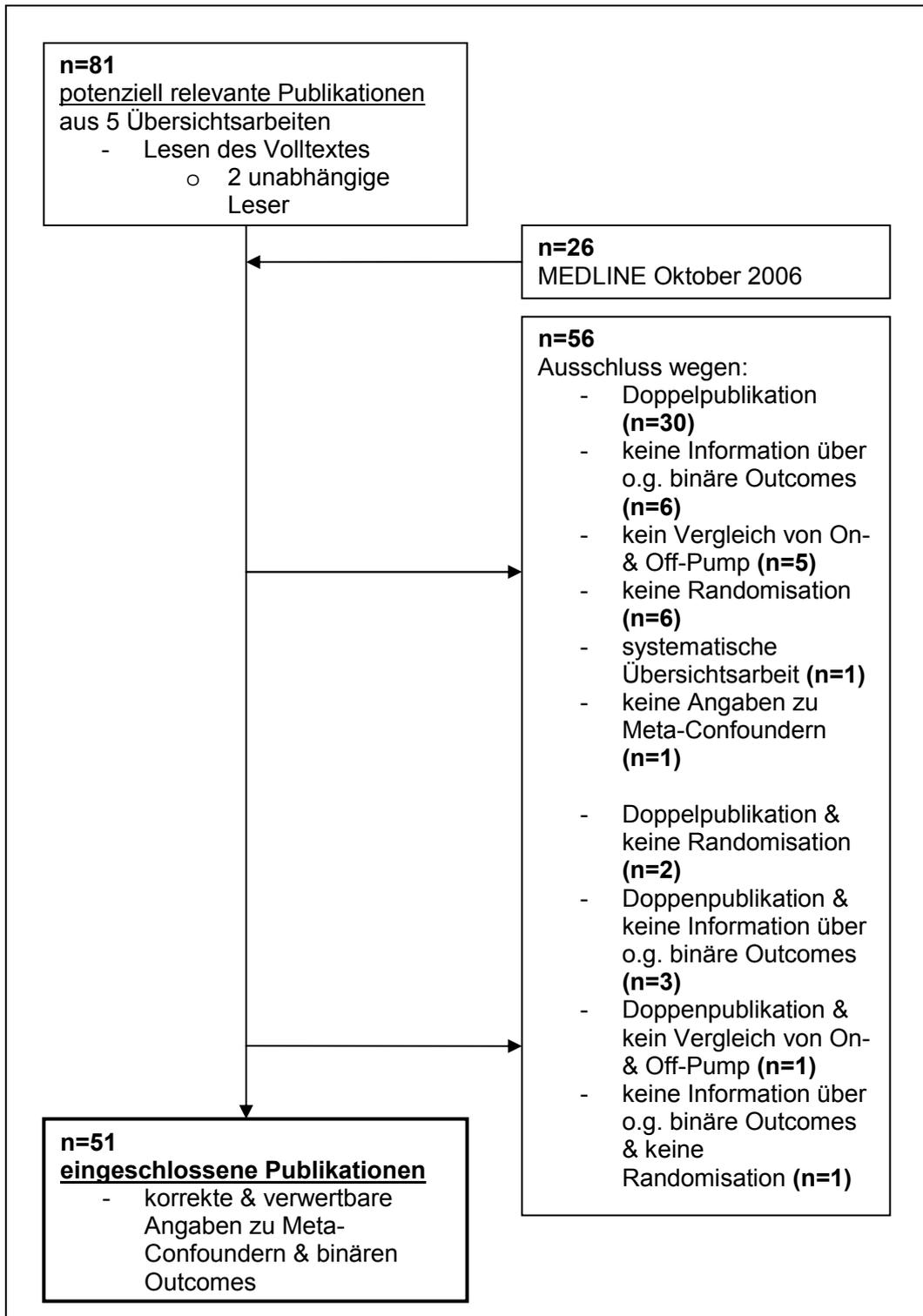


Abbildung 11: Flussdiagramm zur Selektion der RCTs

### 4.1.2 Einschluss der Propensity score Analysen

Im Rahmen einer Metaanalyse von Kuss O., von Salviati B. und Börgermann J.<sup>[54]</sup> wurden 35 aus primär 58 potenziell relevanten PS-Analysen zum Thema gefunden und deren Daten extrahiert. Ein Ausschluss erfolgte zum einen aus methodischen Gesichtspunkten und zum anderen aus Gründen einer insuffizienten Beschreibung sowie Auflistung der Meta-Confounder und der Studienergebnisse (vgl. Abbildung 12).

Von den 58 Studien wurden fünf ausgeschlossen, weil es sich bei den Therapiegruppen nicht um eine On-Pump- bzw. eine Off-Pump-Gruppe handelte. In sechs Analysen wurde die Propensity-score-Methode falsch angewendet, vier der 58 Publikationen erwiesen sich als reine systematische Übersichtsarbeiten ohne Angabe neuerer, primärer Daten und eine als Doppelpublikation. Sechs weitere Studien gingen auf keines der für diese Meta-Propensity-Score-Analyse relevanten Outcomes ein, während eine Studie ihre Resultate lediglich in erzählerischer Form ohne genauere Darlegung von numerischen Werten darstellte und somit ausgeschlossen wurde. Von den 35 übrigen Propensity-score-Analysen konnten zehn aufgrund fehlender Aussagen zu den teilnehmenden Probanden für diese Arbeit nicht genutzt werden.

Im Oktober 2006 wurden nach erneuter MEDLINE-Suche sechs weitere PS-Analysen gefunden. Bei einer dieser sechs Analysen handelte es sich um eine Doppelpublikation, eine Arbeit verglich keine Patienten, die einer CABG bzw. OPCAB unterzogen wurden und aus einer waren keine Daten zu mindestens eines der zehn oben angegebenen dichotomen Outcomes zu extrahieren.

Das nachfolgende Flussdiagramm (Abbildung 12) stellt den Prozess der Studienselektion graphisch dar.

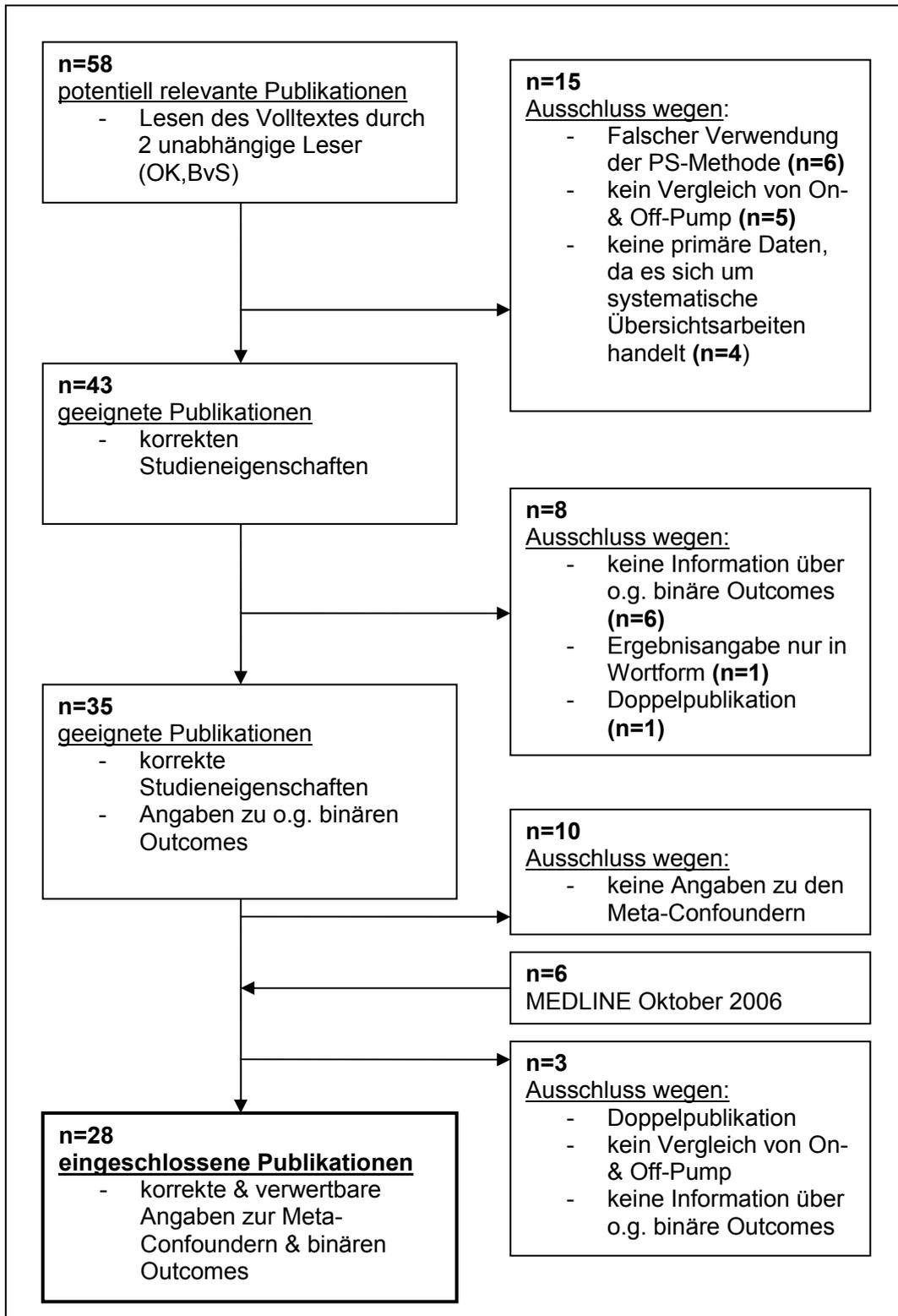


Abbildung 12: Flussdiagramm zur Selektion der PS-Analyse

### 4.2 Verteilung der Meta-Confounder vor Meta-Matching

Unter Einhaltung der Festlegung, dass die stetigen Meta-Confounder in mindestens zwei Drittel, sowohl aller RCTs als auch PS-Analysen enthalten sein müssen, konnten die jeweiligen Meta-Propensity-scores auf der Basis von zwei kategorischen und fünf stetigen Meta-Confoundern berechnet werden (vgl. Tabelle 1). Bei den stetigen Meta-Confounder war keiner in allen Studien vertreten. Dazu gehören das durchschnittliche Lebensalter, der Anteil männlicher Patienten, früher erlittene Myokardinfarkte, diabetisch Erkrankter und die linksventrikuläre Ejektionsfraktion. Es zeigen sich vor dem Meta-Matching bezüglich der kategorischen und stetigen Meta-Confounder zwischen den Studienformen beträchtliche Unterschiede. Wie zu erwarten und eingangs erwähnt, ist die Population in den PS-Analysen mit durchschnittlich 65,8 Lebensjahren älter als die der RCTs mit 63,1 Lebensjahren. Auch die vermeintlich höhere kardiale Krankheitsausprägung im Sinne der geringeren durchschnittlichen LVEF, 58,8% im Vergleich zu 62,7% und der höheren Rate stattgehabter MI, 44,5% zu 41,6% sowie der größere Anteil an Komorbidität bzw. Risikofaktoren, hier repräsentiert durch an Diabetes mellitus Leidender, mit 26,2% zu 24,4%, liegt auf Seiten der PS-Analysen. In Bezug auf die Geschlechtsverteilung wurden im Rahmen der randomisierten Studien mit 77,1% statt 72,1% mehr männliche Probanden beobachtet als bei den PS-Analysen.

Die kategorischen Meta-Confounder waren in allen eingeschlossen randomisierten und nicht randomisierten Studien zu finden. Die PS-Analysen wurden mit 36% (zehn von 28 PS-Analysen) sowie 32% verhältnismäßig häufiger in Nordamerika und multizentrisch praktiziert als die RCTs mit 10% und 6%. Die Durchführungen fanden jedoch am häufigsten, 61% der nRCTs und 71% der RCT, in Europa statt.

Die in Tabelle 1 den jeweiligen Meta-Confoundern zugeordneten p-Werte verdeutlichen den starken und signifikanten Unterschied der Eigenschaften beider Studienformen, vor allem bei der Studienregion, der Anzahl der beteiligten Zentren und dem mittleren Patientenalter, sodass ein Vergleich von RCTs und PS-Analysen unter diesen Umständen kein valides Ergebnis produzieren könnte. Die gegebenen standardisierten Differenzen vervollständigen diese Tatsache aufgrund des Kriteriums, dass jeder gegebene Wert der die 10%-Grenze überschreitet eine Annahme als balancierten Verteilung der stetigen Meta-Confounder per definitionem nicht erlaubt <sup>[57]</sup>.

**Tabelle 1: Verteilung der Meta-Confounder in den Studiengruppen für alle eingeschlossenen Studien vor dem Meta-Matching**

Meta-Confounder	PS-Analysen (n=28)	RCTs (n=51)	p-Wert	SD
<b>Kategorische Meta-Confounder:</b>				
<b>Studienregion</b>			<b>0.007</b>	
Europe	17 (61%)	36 (71%)		
Nordamerika	10 (36%)	5 (10%)		
Andere	1 (3%)	10 (19%)		
<b>Zenteranzahl</b>			<b>0.006</b>	
1	18 (65%)	47 (92%)		
>1	9 (32%)	3 (6%)		
keine Angabe	1 (3%)	1 (2%)		
<b>Kontinuierliche Meta-Confounder:</b> (Durchschnittswerte)				
Alter (in a)	65.8 (58.5 – 73.0, 27)	63.1 (48.3 – 75.5, 50)	<b>0.002</b>	<b>75.1%</b>
Männl. Patienten (in %)	72.1 (0.0 – 90.4, 25)	77.1 (25.0 – 89.2, 48)	0.138	<b>-37.0%</b>
Präop. MI (>1 Monat, in %)	44.5 (11.6 – 68.0, 19)	41.6 (8.3 – 73.3, 29)	0.480	<b>21.0%</b>
LVEF (in %)	58.8 (36.4 – 67.3, 23)	62.7 (44.5 – 75.0, 44)	<b>0.033</b>	<b>-55.9%</b>
Diabetes mellitus (in %)	26.2 (15.2 – 46.8, 25)	24.4 (0.0 – 52.7, 36)	0.595	<b>13.9%</b>

**kontinuierliche Meta-Confounder (kont. MC) = absoluter Wert (Verteilung, gegeben in n Studien)**

### **4.3 Verteilung der Meta-Confounder nach Meta-Matching**

Nach dem anhand der errechneten, studieneigenen Meta-Propensity-scores bzw. deren korrespondierende linearen Meta-Prädiktoren praktiziertem Meta-Matching sind auf Seiten der PS-Analysen weitere 18 und der RCTs zusätzliche 22 aus dem Vergleich der Studienformen ausgeschlossen worden, da ihnen keine Meta-PS-kompatiblen Studien zugeordnet werden konnten. Es sind somit 10 bzw. 29 der ursprünglich vor Wirksamwerden der Ein- und Ausschlusskriterien, sowie der Meta-Matching-Prozedur insgesamt vorhandenen 171 Publikationen in den Vergleich eingegangen. Auf Patientenebene sind Daten von 25.552 und 2.723 Probanden, die im Rahmen von PS-Analysen und randomisierten Studien postoperativ bei unterzogenen CABG und OCABG beobachtet wurden, eingeflossen. Davon waren 8.584 Patienten, 7.242 in den RCTs und 1.342 in den PS-Analysen, in einer OPCAB-Gruppe.

Das Bild der gelisteten Meta-Confounder hat sich nach dem Meta-Matching dahingehend geändert, dass der Unterschied zwischen den Studien- und Patienteneigenschaften minimiert bzw. in Anbetracht der Studienregion sogar aufgehoben wurde. Dass es keinen signifikanten Unterschied zwischen den Studien- und Patientenmerkmalen gibt, zeigen die p-Werte und standardisierten Differenzen, welche in Tabelle 2 aufgelistet sind. Drei der fünf stetigen Meta-Confounder weisen standardisierte Differenzen von unter 10% auf. Weiterhin bestehende moderate Dysbalancen zwischen beiden Studienformen sind lediglich bezüglich des mittleren Anteils männlicher Patienten und präoperativer Myokardinfarkte festzustellen.

**Tabelle 2: Verteilung der Meta-Confounder in den Studiengruppen für alle eingeschlossenen Studien nach dem Meta-Matching**

Meta-Confounder	PS-Analysen (n=10)	RCTs (n=29)	p-Wert	SD
<b>Kategorische Meta-Confounder:</b>				
<b>Studienregion</b>			0.999	
Europe	8 (80%)	23 (80%)		
Nordamerika	1 (10%)	3 (10%)		
Andere	1 (10%)	3 (10%)		
<b>Zenteranzahl</b>			0.631	
1	8 (80%)	25 (86%)		
>1	2 (20%)	3 (10%)		
keine Angabe	0 (0%)	1 (4%)		
<b>Kontinuierliche Meta-Confounder:</b> (Durchschnittswerte)				
Alter (in a)	64.1 (58.5 – 70.1, 10)	63.9 (59.3 – 75.5, 28)	0.916	<b>3.9%</b>
Männl. Patienten (in %)	80.5 (69.5 – 90.4, 9)	76.9 (25.0 – 89.2, 28)	0.431	30.5%
Präop. MI (>1 Monat, in %)	44.0 (17.5 – 68.0, 8)	39.9 (8.3 – 73.3, 16)	0.530	27.6%
LVEF (%)	61.1 (50.6 – 67.3, 9)	60.7 (47.8 – 68.5, 25)	0.861	<b>6.9%</b>
Diabetes mellitus (in %)	24.8 (15.2 – 40.0, 10)	24.4 (12.0 – 52.7, 18)	0.925	<b>-3.7%</b>

**kontinuierliche Meta-Confounder (kont. MC)= absoluter Wert (Verteilung; gegeben in n Studien)**

#### 4.4 Verteilung der Meta-Confounder zwischen den ungematchten Studien

Um die Wirksamkeit der Unterschiedsminimierung bzw. der „nicht meta-randomisierten“ Gleichverteilung von Charakteristika infolge von Meta-PS-Berechnung und Meta-Matching effektiv zu zeigen, sollen im Folgenden neben den in Tabelle 2 dargestellten Meta-Confoundern auch die Meta-Confounder der Studien präsentiert werden, denen aufgrund zu stark divergierender Meta-Propensity-scores keine Partner der jeweilig anderen Gruppe zugeordnet werden konnten. Dabei handelte es sich um 18 PS-Analysen und 22 randomisierte, kontrollierte, klinische Studien, mit 71.926 und 2.235 beobachteten Patienten. Es wurden bei insgesamt 31.469 Patienten, diesbezüglich 30.345 den PS-Analysen und 1.124 den RCTs zugehörig, Bypassoperationen ohne Anwendung der Off-Pump-Technik durchgeführt.

## Ergebnisse

Die p-Werte beider kategorischer Meta-Confunder, des mittleren Lebensalters und der mittleren linksventrikulären Auswurffraktion zeigten einen signifikanten Unterschied zwischen den Studienformen an. Auch die Betrachtung der Standarddifferenzen zeigt im Vergleich zur Situation vor dem Meta-Matching und insbesondere zu den gematchten Studien eine viel höhere Dysbalance in der Merkmalsverteilung an. Lediglich in Bezug auf den Anteil präoperativ erlittenen Myokardinfarkte liegt ein balanciertes Verhältnis vor. Dieser Meta-Confunder zeigt in den gematchten und ungematchten Gruppen ein dem eigentlichen Ziel von Meta-Propensity-Score-Berechnung und Meta-Matching entgegen gesetztes Verhalten.

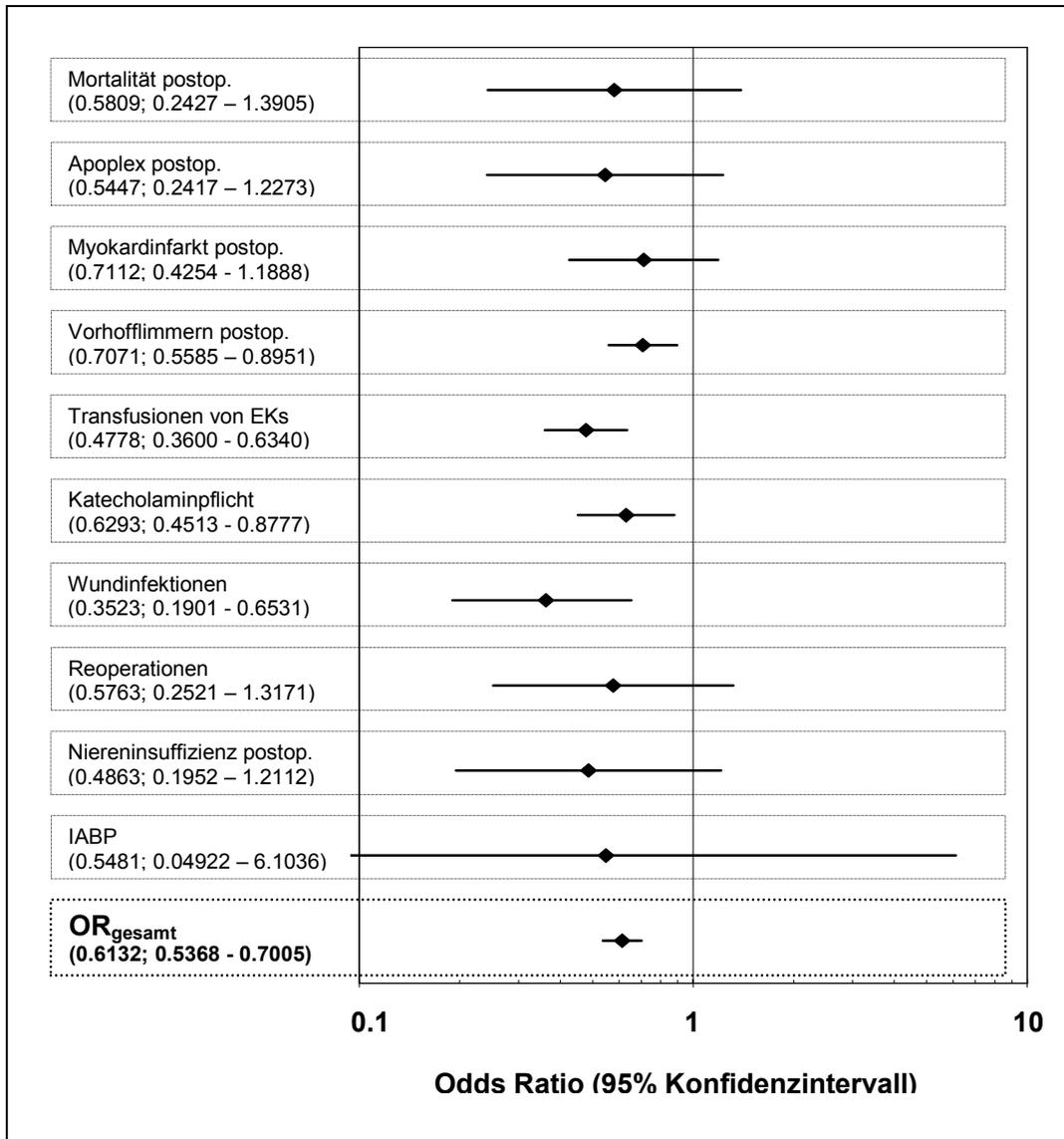
**Tabelle 3: Verteilung der Meta-Confunder in den Studiengruppen für alle ungematchten Studien**

Meta-Confunder	PS-Analysen (n=18)	RCTs (n=22)	p-Wert	SD
<b>Kategorische Meta-Confunder:</b>				
<b>Studienregion</b>			<b>0.003</b>	
Europe	9 (50%)	13 (59%)		
Nordamerika	9 (50%)	2 (9%)		
Andere	0 (0%)	7 (32%)		
<b>Zenteranzahl</b>			<b>0.002</b>	
1	10 (56%)	22 (100%)		
>1	7 (39%)	0 (0%)		
keine Angabe	1 (5%)	0 (0%)		
<b>Kontinuierliche Meta-Confunder:</b> (Durchschnittswerte)				
Alter (in a)	66.9 (62.0 – 73.0, 17)	62.0 (48.3 – 67.8, 22)	<b>&lt;0.001</b>	<b>126.8%</b>
Männl. Patienten (in %)	67.4 (0 – 83.2, 16)	77.4 (51.7 – 88.0, 20)	0.048	<b>-68.7%</b>
Frühere MI (>1 Monat, in %)	44.9 (11.6 – 63.8, 11)	43.8 (25.9 – 67.1, 13)	0.835	8.7%
LVEF (in %)	57.1 (36.4 – 64.8, 14)	65.3 (44.5 – 75.0, 19)	<b>0.004</b>	<b>-108.8%</b>
Diabetes mellitus (in %)	27.0 (15.7 – 46.8, 15)	23.7 (0.0 – 52.0, 18)	0.504	<b>23.6%</b>

**kont. MC = absoluter Wert(Verteilung, gegeben in n Studien)**

## 4.5 Klinische Zielgrößen

### 4.5.1 Effektschätzer der klinischen Zielgrößen



**Abbildung 13: Meta-Odds Ratios der binären Outcomes aller gematchten RCTs (n=29); [OR(95%-CI)]**

In Abbildung 13 sind die Meta-Odds-Ratios der binären Outcomes aller gematchten randomisierten Studien dargestellt. Bezüglich der Gesamtheit dieser Outcomes gibt es einen signifikant besseren Therapieeffekt auf Seiten der OPCAB-Prozedur. Jedoch im Vergleich zu den Ergebnisse der zehn gematchten PS-Analysen (Abbildung 14) zeigen sich, einzeln betrachtet, bei den dokumentierten postoperativen Mortalitäten, Schlaganfällen, Myokardinfarkten, blutungsbedingten Revisionsoperationen, akuten Niereninsuffizienzen und Einsätzen des IABP keine signifikanten Differenzen in den Meta-Odds-Ratios beider untersuchter Gruppen.

## Ergebnisse

Die Darstellung der Meta-Odds-Ratios aller PS-Analysen, denen Meta-PS-kompatible RCTs zugeordnet werden konnten, zeigt ebenso, dass in der Gesamtheit aller klinischen Outcomes die Off-Pump-Technik einen signifikant besseren Therapieeffekt vorweisen kann als die konventionelle Methode. In Bezug auf den beobachteten Anteil postoperative eingetretener Myokardinfarkte und Vorhofflimmern sowie der Notwendigkeit eines IABP zur hämodynamischen Unterstützung konnte in dieser Gruppe kein signifikanter Unterschied festgestellt werden.

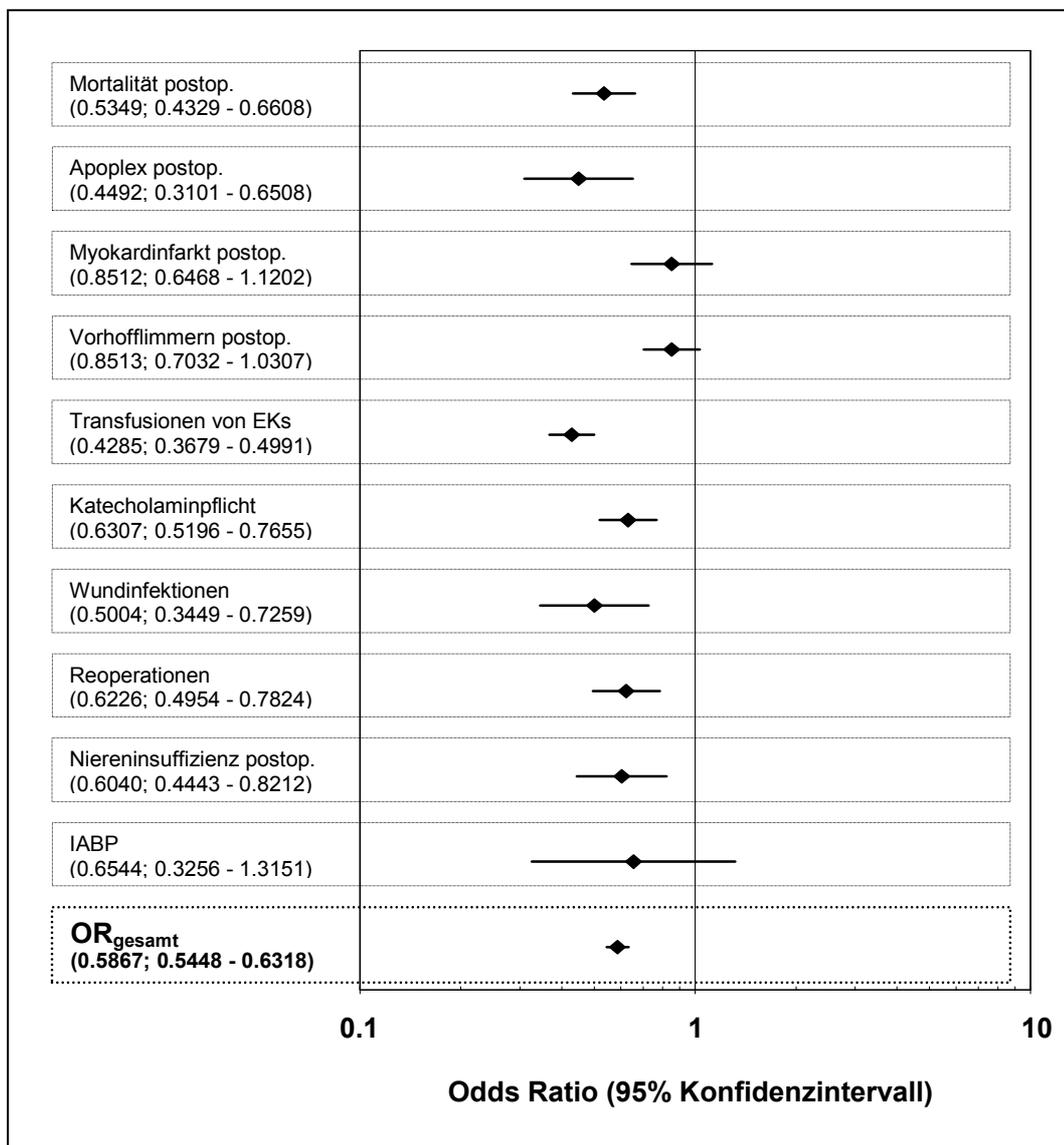


Abbildung 14: Meta-Odds Ratio der binären Outcomes aller gematchten PS-Analysen (n=10); [OR(95%-CI)]

4.5.2 Differenzen der Meta-Odds-Ratios

Die Differenzen der in Abbildung 13 und 14 dargestellten Meta-Odds-Ratios lassen in Abbildung 15 erkennen, dass der Vergleich von randomisierten, kontrollierten, klinischen Studien und Propensity-score-Analysen keinen signifikanten Unterschied weder in der Gesamtheit, noch in der Einzelaufführung der Ergebnisse beider Studienformen aufzeigen kann. Die Differenz der  $OR_{Gesamt}$  ist -0,03 mit einem 95%-Konfidenzintervall von -0,12 bis 0,07, welches die Differenz Null einschließt. Alle Konfidenzintervalle der jeweiligen Outcomes schließen die Differenz 0 ein.

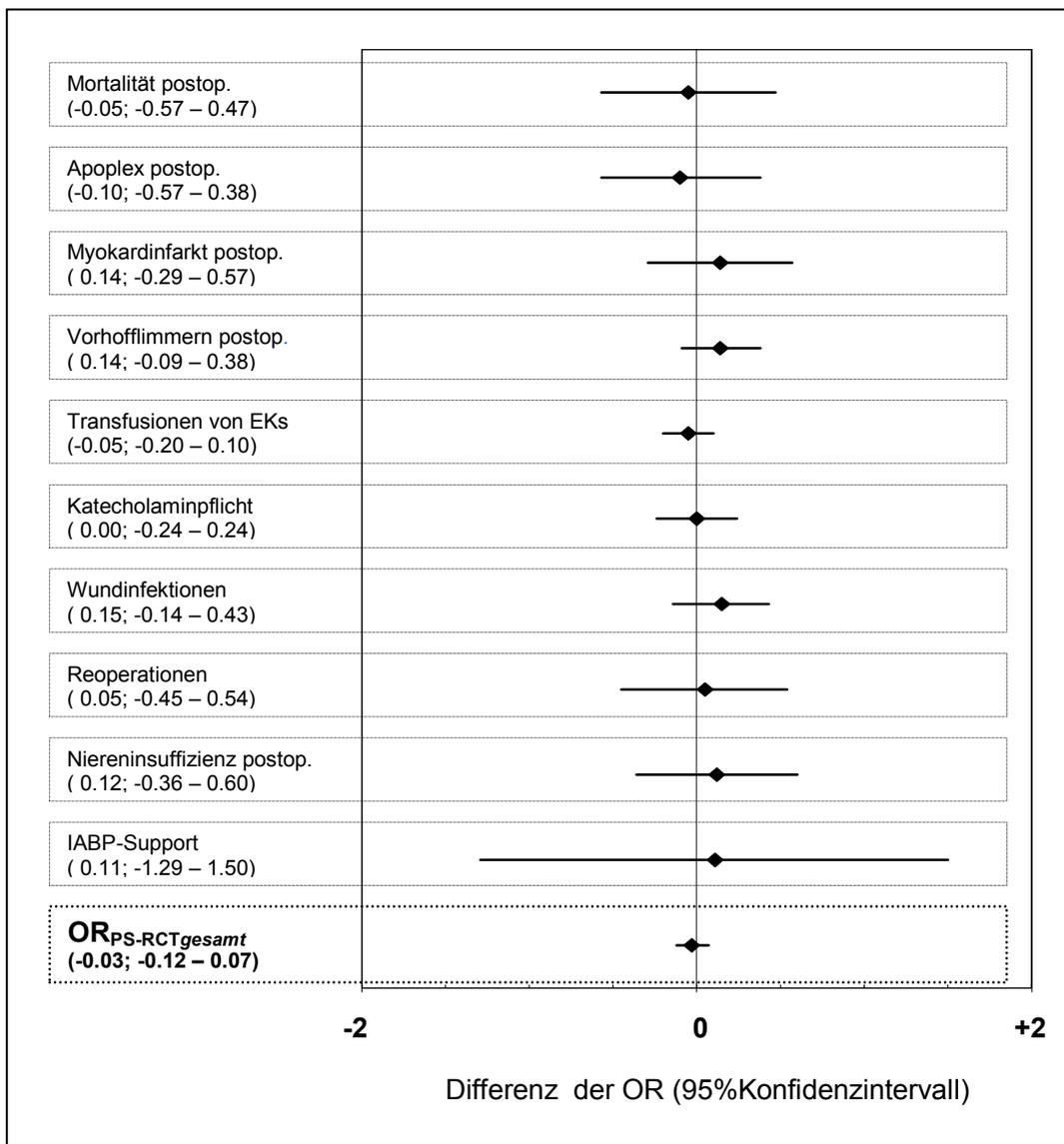
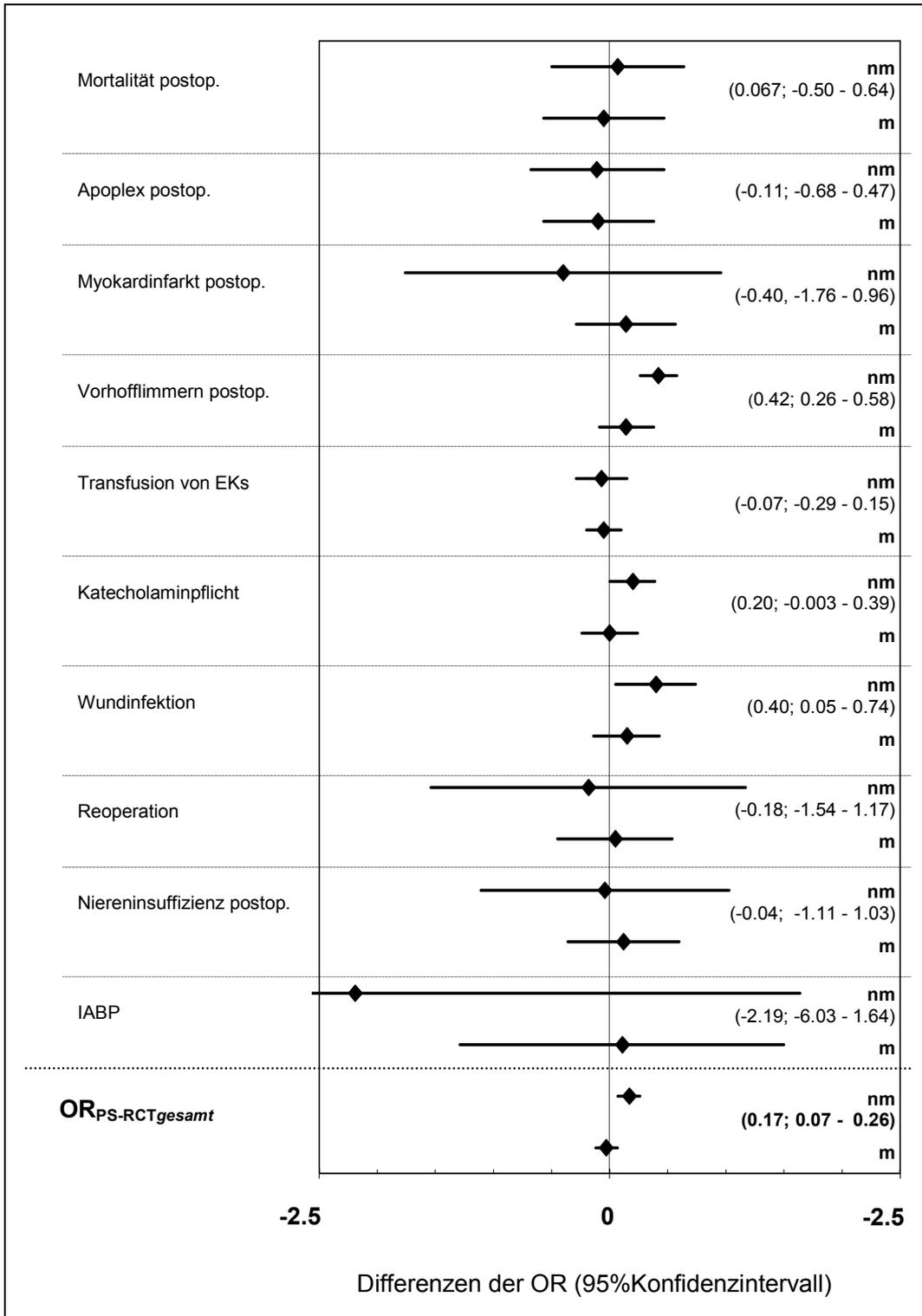


Abbildung 15: Differenzen der Odds Ratios beider Studientypen im gematchten Sample; [OR<sub>PS-RCT</sub>; 95%-CI]

## Ergebnisse

In Abbildung 16 wird die Auswirkung der Meta-Confounder-Adjustierung und des Meta-Matchings auf die Vergleichsergebnisse verdeutlicht.



**Abbildung 16: Gegenüberstellung aller Differenzen der Meta-Odds-Ratios bezüglich gematchter (m) und ungematchter (nm) Studien; [OR<sub>PS-RCT</sub>; 95%-CI]<sub>ungematchtes Sample</sub>**

Die zusammengefassten Differenzen der Meta-Odds-Ratios verzeichnen im ungematchten Sample (nm) einen signifikanten Ergebnisunterschied zwischen den Studienformen. Diese Divergenz der Outcomes lässt sich nach erfolgtem Meta-Matching nicht mehr darstellen. Auch die Ergebnisunterschiede in den Einzelaufführungen, postoperatives Vorhofflimmern und Katecholaminpflicht, sind im gematchten Sample (m) nicht vorhanden.

## 5. Diskussion

Die mittels RCT und nRCT gemessenen Therapieeffekte von On-Pump und Off-Pump-Operationen in der Bypasschirurgie sind sehr ähnlich in einer meta-gematchten Auswahl von Studien. Dieses Ergebnis stimmt mit Resultaten vieler kürzlich durchgeführter systematischer Vergleiche von randomisierten, kontrollierten, klinischen Studien und nicht randomisierten, klinischen Studien überein [59; 60; 61; 62; 63; 64; 65; 66]. Im Rahmen des hier durchgeführten Vergleiches von RCTs und PS-Analysen konnten die von Deeks et al. [36] genannten und sich auf eine derartige Gegenüberstellung der Studienformen erschwerend und verzerrend auswirkenden Punkte kontrolliert und vermieden werden. Bei diesen Punkten handelt es sich vor allem um Unterschiede zwischen den Studienformen bezüglich der Studienpopulation, der Therapie, der Kontrollbehandlung, der Dauer der Beobachtung, der zeitlichen Beobachtungsintervalle und der Studienergebnisse.

Durch die Meta-Propensity-score-Analyse konnte eine Gleichheit zwischen den Populationen erreicht werden. Des Weiteren wurden Unterschiede zwischen den einzelnen nicht randomisierten Studien in Bezug auf ihr Analysedesign aufgehoben, weil lediglich die PS-Analyse zur Confounderadjustierung verwendet wurde. Außerdem hat die Tatsache bestanden, dass randomisierte und nicht randomisierte Studien im vorliegenden Fall identische Interventionen und Kontrolltherapien, OPCAB versus CABG, untersuchten. Es wurden nur die Ergebnisse, die innerhalb des ersten postoperativen Monats dokumentiert wurden, miteinander verglichen. Somit lagen identische Beobachtungszeiträume vor. Zusätzlich gab es Überschneidungen in den zeitlichen Intervallen, in denen RCTs und PS-Analysen durchgeführt wurden. Das bedeutet, dass die medizinischen Standards, die Operationen an sich sowie die perioperative Patientenbetreuung betreffend, bezüglich des medizinischen Erkenntnisstandes in den Studien beider Formen vergleichbar sind.

Des Weiteren berichteten Deeks et al. in ihrer Ausführung [36] über die Problematik von Dosisunterschieden. Damit sind sich unterscheidende, nicht standardisierte Therapiedosierungen gemeint, welche verschiedene Outcomes produzieren und vor allem die interne Validität eines Therapie- aber auch Studienvergleichs senken können. Da es sich im Fall dieser Arbeit aber um operative Eingriffe handelt, gibt es lediglich binäre „Dosisoptionen“. Es gab lediglich die Möglichkeit, mit Off-Pump-Technik oder konventionellem Einsatz

einer Herzlungenmaschine, also am schlagenden oder kardioplegischen Herzen zu operieren. Außerdem sind die aus den RCTs und nRCTs extrahierten binären Outcomes, wie Mortalitäts- und Vorhofflimmerrate identisch. Insbesondere diese Outcomes lassen einen sehr validen Vergleich zu, weil es zwischen den einzelnen Publikationen in Bezug auf die Definition dieser Outcomes keine Unstimmigkeiten geben kann.

Aus den genannten Gründen lassen sich signifikante oder nicht signifikante, vorhandene bzw. nichtvorhandene Abweichungen mit hoher interner Validität direkt darstellen und bewerten.

Die Hauptproblematik eines derartigen Vergleiches ist das Publikationsbias. Tatsächlich publizierte Artikel sind zumeist eine verzerrte Stichprobe aller Forschungsergebnisse, denn es werden wahrscheinlich mit überwiegendem Anteil nur Studien publiziert, in denen ein signifikanter positiver Therapieeffekt festgestellt wird <sup>[13]</sup>. In dieser Arbeit gibt es zwei Ebenen von Publikationsbias. Zum einen können die Originalstudien nicht veröffentlicht worden sein oder zum anderen besteht die Möglichkeit, dass bereits publizierte Studien nicht in die Übersichtsarbeiten von Cheng et al. <sup>[49]</sup>, Raja <sup>[50]</sup>, Reston et al. <sup>[51]</sup>, Sedrakyan et al. <sup>[52]</sup> oder Wijeyesundera et al. <sup>[53]</sup> eingeschlossen wurden. Im Falle der ersten Situation ist nicht bekannt, um wie viele nicht publizierte Studien es sich auf Seiten der RCTs und auf Seiten der Propensity-score-Analysen handelt bzw. in welcher Richtung, positiver oder negativer, die vermeintlichen, nicht publizierten Ergebnisse ausfallend, den hier durchgeführten Vergleich und die damit erhaltenen Ergebnisse beeinflusst hätten. Aufgrund der erneuten MEDLINE-Studiensuche im Oktober 2006 konnte die zweite Ebene des Publikationsbias vermieden werden.

Eine weitere Gefahr für eine Verzerrung der Ergebnisse dieser Arbeit besteht durch die gegebenenfalls unterschiedlichen Outcomes zwischen Intention-to-treat-Studien<sup>14</sup> (ITT) und explanativen Studien<sup>15</sup> (Per-Protocoll-Analyse), da bei den RCTs häufige Cross-over<sup>16</sup> beobachtet wurden. In der vorliegenden Arbeit wurden, wenn beide Auswertungsmethoden angegeben waren, die ITT

---

<sup>14</sup> Das Prinzip der Intention-to-treat-Studien ist die Ergebnisanalyse entsprechend der zugewiesenen (randomisierten) Therapien, auch wenn während der Studienphase ein Wechsel in die andere Gruppe stattfand <sup>[13]</sup>.

<sup>15</sup> Das Prinzip einer explanativen Studie ist die Per-Protocoll-Analyse. Die Ergebnisanalyse erfolgt entsprechend der tatsächlich durchgeführten Therapie <sup>[13]</sup>.

<sup>16</sup> Cross-over beschreibt den Wechsel in eine andere Therapiegruppe. Dabei kann es sich um eine Wechsel in die Beobachtungs- oder Kontrollgruppe handeln <sup>[13]</sup>.

bevorzugt. Der Grund dafür ist der Vergleich von PS-Analysen mit randomisierten, kontrollierten, klinischen Studien. Verwendet man die Per-Protocoll-Analyse hebt man theoretisch die Randomisation auf, indem man entsprechend der erhaltenen Behandlung analysiert <sup>[13]</sup>. Folglich wäre die per protocoll ausgewertete Studie per definitionem keine RCT mehr, sondern eine Kohortenstudie und für die Bearbeitung der Fragestellung dieser Arbeit nicht brauchbar.

Unbekannte Confounder stellen bei nicht randomisierten Studien ein Problem dar, weil die zu messende Prädiktor-Outcome-Beziehung für diese nicht adjustiert werden kann. Auch in die Propensity-score-Berechnung können unbekannte Merkmale nicht mit einfließen. Es muss in Betracht gezogen werden, dass bei einer Meta-Propensity-score-Analyse unbekannte Meta-Confounder unentdeckt bleiben und trotz wirksamer Adjustierung nicht alle Unterschiede zwischen den RCTs und PS-Analysen auszugleichen sind. Diese unbekanntes Meta-Confounder können weiterhin Einfluss auf die Ergebnisse dieser Arbeit ausüben. Eine Meta-Randomisation kann die Kontrolle dieser unbekanntes Variablen verbessern. Ein nicht zu messender Meta-Confounder ist beispielsweise die Zustimmung zur Teilnahme an einer Studie. Patienten, die in eine Teilnahme an einer RCT einwilligen würden, unterscheiden sich annehmbar in ihrem Charakter, demzufolge in ihrer Lebensweise und in Bezug auf gesundheitliche Risikofaktoren (Meta-Confounder) von denen, die eine derartige Teilnahme völlig ausschließen. Wie genau dieser Unterschied aussehen würde, ist nicht bekannt. Es besteht die Tatsache, dass an einer RCT teilnehmende Probanden ihr Einverständnis geben und Patienten zur Teilnahme an einer nicht randomisierten Studie nicht gefragt werden müssen. Es ist unsicher, wie viele in einer nRCT eingeschlossenen Testpersonen überhaupt Ihr Einverständnis geben würden. Also ein so verborgen gebliebener Unterschied in den Studienpopulationen kann weiterhin die Ergebnisse dieser Arbeit beeinflussen.

Neben den Outcomes Mortalität und Vorhofflimmern wurden Outcomes verglichen, die in ihrer Definition Schwankungen unterliegen können. Es war z. B. in den Studien nicht ersichtlich, ab welchem Hämoglobinwert bzw. Hämatokrit Erythrozytenkonzentrate transfundiert wurden bzw. wie die entsprechenden Kriterien zu welchem Ausmaß erfüllt sein mussten, damit ein akutes Nierenversagen, ein Apoplex, eine Wundinfektion oder ein Myokardinfarkt vorliegen bzw. die Notwendigkeit eines IABP, von Katecholaminen sowie einer

Revisionsoperation besteht. Es ist ungewiss, ob diese Outcomes wirklich valide gegenübergestellt werden können.

Wie in der Einleitung dieser Arbeit beschrieben, führt eine kleinere Studienpopulation zu einer Risikoanhebung für das Auftreten eines  $\beta$ -Fehlers und somit zu einer Verringerung der statistischen Trennschärfe. Diese Tatsache besteht in dem Fall des hier durchgeführten Vergleiches auch auf Meta-Ebene. Nach dem Meta-Matching ist nur eine geringe Anzahl an RCT und nRCT in den Vergleich eingegangen. Ein tatsächlich vorhandener Unterschied in den Outcomes randomisierter und nicht randomisierter Studien kann also möglicherweise unentdeckt geblieben sein.

Die Ergebnisse dieser Arbeit zeigen, dass im vorliegenden Vergleich keine signifikanten Ergebnisunterschiede zwischen den Studienformen bestehen. Dieses Resultat steht im Widerspruch zur vorherrschenden Meinung, dass Resultate von Therapievergleichen aus nicht randomisierten Studien „aufgeblähte“ positive Therapieeffekte suggerieren [7; 9; 67; 68; 69; 70; 71]. Kunz R. und Oxman A.D. stützen diesen Standpunkt in ihrer Übersichtsarbeit (Kunz R, Oxman AD 1998) [38]. Bei dem Vergleich von randomisierten und nicht randomisierten Studien mit Gegenüberstellungen identischer Interventionen wurden acht Übersichtsarbeiten aufgeführt. Fünf dieser Publikationen beschrieben ein Überschätzen der Therapieeffekte durch nRCT. Bei diesen nRCT handelte es sich vor allem um historische Vergleiche. In zwei Fällen wurde der untersuchte Therapieeffekt durch die nicht randomisierten Studien unterschätzt. Es sei anzumerken dass in einer der beiden Übersichtsarbeiten die Patienten der nRCTs generell eine schlechtere Prognose in Anbetracht eines Therapieerfolges vorwiesen. In einer Übersichtsarbeit gab es zwischen den Ergebnissen aus RCT und nRCT keinen Unterschied.

Bestünde eine generelle Überschätzung von Therapieeffekten infolge der Durchführung einer nRCT und die Möglichkeit, Regeln und Ausmaße der Überschätzung festzulegen, könnte man die Ergebnisse den angeblich wahren Werten anpassen.

Dass diese Überbewertung nicht zutrifft, konnten Concato et al. [7] (s.o.) und Benson et al. [9] in ihren Vergleichen von randomisierten und nicht-randomisierten Studien zeigen. Die Autoren untersuchten Veröffentlichungen beider Studienformen, die identische klinische Fragestellungen bearbeiteten. Benson et al. konnten anhand von 19 Interventionsvergleichen zeigen, dass die kombinierten

Ergebnisse aus 17 der 19 nRCTs innerhalb des 95%-Konfidenzintervalls der RCTs liegen. Dabei war keine Tendenz ersichtlich, ob die Therapieeffekte in den nRCTs positiver oder negativer ausfielen. Die Voraussetzungen für diese Ergebnisse waren kongruente Durchführungsprinzipien, wie identische Ein- und Ausschlusskriterien sowie ähnliche oder gar identische Patientenmerkmale in den Vergleichsgruppen, entsprechend der Empfehlung von Deeks et al.. Folglich besteht für jeden Probanden ein analoges Basisrisiko für ein bestimmtes Outcome (z.B. Tod, Myokardinfarkt, etc.) und somit eine Anpassung an das Format randomisierter, kontrollierter, klinischer Studien. Die Untersuchungsmethode an sich ist somit für eine Wertverschiebung oder -abweichung in eine bestimmte Richtung nicht verantwortlich <sup>[7]</sup>.

## 6. Zusammenfassung

Es konnte anhand der in dieser Arbeit verwendeten klinischen Fragestellung gezeigt werden, dass es keinen signifikanten Unterschied zwischen den aus randomisierten, kontrollierten, klinischen Studien und Propensity-score-Analysen hervorgegangenen Ergebnissen gibt. Bei der jeweiligen Studienform besteht keine Tendenz, Outcomes, die in eine bestimmte Richtung abweichen, zu produzieren, also einen untersuchten Therapieeffekt zu über- bzw. unterschätzen.

Die Propensity-score-Analyse ist nicht nur eine sehr wirksame Methode zur Confounderadjustierung, sondern im vorliegenden Fall unter Betrachtung der internen Ergebnisvalidität eine dem heutigen Qualitätsstandard ebenbürtige Methode zur Evaluation von Therapieeffekten. Es konnte gezeigt werden, dass mittels Propensity score und seinen Anwendungsstrategien, v. a. dem Matching, sehr identische Untersuchungs- und Kontrollgruppen bezüglich der bekannten Confounder gebildet werden können. Die Auswirkung der Gruppenhomogenität auf die erzeugten Outcomes wurde bei der Gegenüberstellung der gematchten und der ungematchten Samples (Abbildung 16) veranschaulicht. Eine geringere Gruppenhomogenität hat eine höhere Ergebnisheterogenität zur Folge.

Im Rahmen dieser Arbeit wurden randomisierte, kontrollierte, klinische Studien und Propensity-score-Analysen am Beispiel operativer Therapiemethoden miteinander verglichen. Es sollten weitere Vergleiche, auch bei der Evaluation von Therapieeffekten medikamentöser Behandlungsstrategien durchgeführt werden.

Jede Studienform hat sowohl ihre Vor- als auch Nachteile. Die Verwendung der jeweiligen Methode muss immer entsprechend der klinischen Fragestellung sowie der praktischen und ethischen Umsetzbarkeit angepasst sein. Die RCTs stellen nach wie vor den Qualitätsstandard zur Evaluation von Therapieeffekten dar.

Abschließend ist festzuhalten, dass die Ergebnisse dieser Arbeit durch eine Vielzahl weiterer derartiger Vergleiche, wenn möglich in meta-randomisierter Form, bestätigt werden müssen, um definitive Schlussfolgerungen hinsichtlich der internen Validität von Propensity-score-Analysen verglichen mit dem heutigen „Goldstandard“ ziehen zu können.

## 7. Literaturverzeichnis

1. Sterne JA, Juni P, Schulz KF et al. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med.* 2002; 21: 1513-1524.
2. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ.* 1996; 312: 1215-1218.
3. 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.
4. Abel U, Koch A. The role randomization in clinical studies: myths and beliefs. *J. Clin. Epidemiol:* 1999; 52: 487 - 97.
5. Adamina M, Guller u, Weber WP, Oertli D. Propensity scores and the surgeon. *Brit. J. Surg.* 2006; 93: 389 - 94.
6. Byar DP, Simon RM, Friedewald WT, et al. Randomized clinical trial: perspective on some recent ideas. *Nengl. J. Med.* 1976; 295: 74 - 80.
7. Concato J, Nirav S, Horwitz RI. Randomized, controlled trials, observational studies and the hierarchy of research designs. *NEngl. J. Med.* 2000; 342: 1887 - 92.
8. Feinstein AR, Current problems and future challenges in randomized clinical trials. *Circulation* 1984; 70: 767-74.
9. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *Nengl. J. Med.* 2000; 342: 1878 - 86.
10. Willich, Stefan N. Randomisierte kontrollierte Studien: Pragmatische Ansätze erforderlich. *Deutsches Ärzteblatt* 103, Ausgabe 39 vom 29.09.2006, Seite A-2524 / B-2185 / C-2107.
11. *Deutsche Zahnärztliche Zeitschrift* 61 (2006) 8, 391-92.

12. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies: a systematic review: *Health Technol Assessment* 1998; 2: i-124.
13. Fletcher RH, Fletcher SW, Wagner EH: *Klinische Epidemiologie, Grundlagen und Anwendung*. Deutschsprachige Ausg. Adaptiert und hrsg. von Johannes Haerting und Christoph Rink. Ullstein Medical, Wiesbaden, 1999: S. 9-13, 17-20, 142, 170-230, 271-273, 368-369.
14. Streptomycin treatment of pulmonary tuberculosis: a Medical Research Council investigation. *BMJ* 1948; 2: 769–82.
15. Rothwell PM. External validity of randomised controlled trials: 'to whom do the results of this trial apply?' *Lancet* 2005; 365(9453): 82-93.
16. Bartlett C, Doyal L, Ebrahim S et al. The causes and effects of socio-demographic exclusions from clinical trials. *Health Technol Assess.* 2005; 9: iii-x, 1.
17. Col NF, Gurwitz JH, Albert JS, Goldberg RJ. Frequency of inclusion of patients with cardiogenic shock in trial of clinical therapy. *Am. J. Cardiol.* 1994; 73: 149 - 57.
18. Heithoff KA, Lohr KN, editors. *Effectiveness and outcomes in health care. Proceedings of an invitational by the institute of Medicine Division of Health Care Services*. Washington D.C.: National Academy Press, 1990.
19. Sackett DL, Haynes RB, Gent M, Taylor DW. Compliance. In: Inman WHW, ed. *Monitoring for drug safety*, Lancaster, UK: MTP Press, 1980.
20. Kaptchuk TJ. The double-blind, randomized, placebo-controlled trial: Gold standard or cold calt? *J. Clin. Epidemio.* 2001; 54: 541 - 49.
21. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. *Evidence based medicine: How to practice and teach EBM*. New York: Churchill Livingtone, 1997.
22. Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika.* 1983; 70: 41-55.

23. Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol* 2003; 158: 280–287.
24. Cook EF, Goldman L. Performance of Tests of Significance Based on Stratification by A Multivariate Confounder Score Or by A Propensity Score. *Journal of Clinical Epidemiology*. 1989; 42: 317-324.
25. Robinson LD, Jewell NP. Some Surprising Results About Covariate Adjustment in Logistic-Regression Models. *International Statistical Review*. 1991; 59: 227-240.
26. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998; 17: 2265-2281.
27. Myers WO, Gersh BJ, Fisher LD, Mock MB, Holmes DR, Schaff HV et al. Time to first new myocardial infarction in patients with mild angina and three-vessel disease comparing medicine and early surgery: a CASS registry study of survival. *Coronary Artery Surgery Study*. *Ann Thorac Surg* 1987; 43: 599–612.
28. Winkelmayr WC, Kuth T. Propensity scores: help or hype? *Nephrol Dial Transplant* 2004; 19: 1671–1673.
29. Braitman LE, Rosenbaum PR. Rare outcomes, common treatments: analytic strategies using propensity scores. *Ann Intern Med* 2002; 137: 693–695.
30. Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat* 1985; 39: 33–38.
31. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American statistical Association* 1984; 79(387): 516-524.
32. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997; 127: 757- 763.

33. Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *J Clin Epidemiol* 2001; 54: 979–985.
34. Pedruzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996; 49: 1373–1379.
35. Calafiore AM, Teodori G, Mezzetti A et al: Intermittent antegrade warm blood cardioplegia. *Ann Thorac Surg* 1995; 59: 398–402.
36. Deeks JJ, Dinnes J, D'Amico R et al. Evaluating non-randomised intervention studies. *Health Technol Assess.* 2003; 7: iii-173.
37. Ioannidis JP, Haidich AB, Pappa M et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA.* 2001; 286: 821-830.
38. Kunz R, Oxman AD. The unpredictable paradox: Review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998; 317: 1185-1190.
39. 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.
40. Osswald BR. Elektive Koronarrevaskularisation: Was ist gesicherte Erkenntnis? *Z Herz- Thorax- Gefäßchir* 2008 22: 70–76.
41. Wahlers T, Wittwer T, Adams DH: *Cardiac Surgical, Operative atlas.* Lehmanns Media, Berlin, 2008, S. 2-44.
42. Klungel OH, Martens EP, Psaty BM et al. Methods to assess intended effects of drug treatment in observational studies are reviewed. *J Clin Epidemiol.* 2004; 57: 1223-1231.
43. Buckberg, GD. Strategies and logic of cardioplegic delivery to prevent, avoid and reverse ischemic and reperfusion damage. *J. Thorac. Cardiovasc. Surg.* 1987; 93(1): 127-39.

44. Schnabel Ph A, Gebhard M M, Pomykaj T. Myocardial Protection: Left Ventricular Ultrastructure after Different Forms of Cardiac Arrest. *Thorac cardiovasc Surg* 1987; 35: 148-156.
45. Gulielmos V: Beating heart bypass surgery and minimally invasive conduit harvesting, *Cardiosurgical techniques, anesthesia managemanet*. Steinkopff Verlag, Darmstadt, 2004, S. 51-68.
46. Rastan AJ, Walther Th, Falk V, Gummert FF, Eckenstein JI, Mohr FW. Off-Pump-Koronarrevaskularisation: "State of the art 2006 und Ergebnisse im Vergleich zur konventionellen Bypassoperation. *Herz* 2006; 31 (5): 384-395.
47. Gummert JF, Funkat A, Krian A. Cardiac surgery in Germany during 2004: a report on behalf of the German Society for Thoracic and Cardiovascular Surgery. *Thorac Cardiovasc Surg* 2005; 53: 391–9.
48. 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.
49. Cheng DC, Bainbridge D, Martin JE et al. Does off-pump coronary artery bypass reduce mortality, morbidity, and resource utilization when compared with conventional coronary artery bypass? A meta-analysis of randomized trials. *Anesthesiology*. 2005; 102: 188-203.
50. Raja SG. Pump or no pump for coronary artery bypass: current best available evidence. *Tex Heart Inst J*. 2005; 32: 489-501.
51. Reston JT, Tregear SJ, Turkelson CM. Meta-analysis of short-term and mid-term outcomes following off-pump coronary artery bypass grafting. *Annals of Thoracic Surgery*. 2003; 76: 1510-1515.
52. Sedrakyan A, Wu AW, Parashar A et al. Off-pump surgery is associated with reduced occurrence of stroke and other morbidity as compared with traditional coronary artery bypass grafting - A meta-analysis of systematically reviewed trials. *Stroke*. 2006; 37: 2759-2769.

53. Wijeyesundera DN, Beattie WS, Djaiani G et al. Off-pump coronary artery surgery for reducing mortality and morbidity: meta-analysis of randomized and observational studies. *J Am Coll Cardiol* 2005; 46: 872-882.
54. 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.
55. Blackstone EH. Comparing apples and oranges. *J. Thorac. Cardiovasc. Surg.* 2002; 123(1): 8 - 15.
56. Joffe MM, Rosenbaum PR. Invited commentary: Propensity scores. *American Journal of Epidemiology* 1999 August 15; 150(4): 327-33.
57. Normand SLT, Landrum NB, Guadagnoli E et al. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: A matched analysis using propensity scores. *Journal of Clinical Epidemiology.* 2001; 54: 387-398.
58. Cepeda MS, Boston R, Farrar JT et al. Optimal matching with a variable number of controls vs. a fixed number of controls for a cohort study: trade-offs. *Journal of Clinical Epidemiology.* 2003; 56: 230-237.
59. Furlan AD, Tomlinson G, Jadad AA, Bombardier C. Methodological quality and homogeneity influenced agreement between randomized trials and nonrandomized studies of the same intervention for back pain. *J Clin Epidemiol* 2008 March; 61(3): 209-31.
60. Hernan MA, Alonso A, Logan R et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology* 2008 November; 19(6): 766-79.
61. King M, Nazareth I, Lampe F et al. Impact of participant and physician intervention preferences on randomized trials: a systematic review. *JAMA* 2005 March 2; 293(9): 1089-99.
62. Noel PH, Larme AC, Meyer J, Marsh G, Correa A, Pugh JA. Patient choice in diabetes education curriculum. Nutritional versus standard content for type 2 diabetes. *Diabetes Care* 1998 June; 21(6): 896-901.

63. Shadish WR, Clark MH, Steiner PM. Can Nonrandomized Experiments Yield Accurate Answers? A Randomized Experiment Comparing Random and Nonrandom Assignments. *Journal of the American Statistical Association* 2008 December; 103(484): 1334-43.
64. Smeeth L, Douglas I, Hall AJ, Hubbard R, Evans S. Effect of statins on a wide range of health outcomes: a cohort study validated by comparison with randomized trials. *Br J Clin Pharmacol* 2009 January; 67(1): 99-109.
65. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. *BMJ* 2009; 338: b81.
66. Vist GE, Hagen KB, Devereaux PJ, Bryant D, Kristoffersen DT, Oxman AD. Systematic review to determine whether participation in a trial influences outcome. *BMJ* 2005 May 21; 330(7501): 1175.
67. Chalmers TC, Celano P, Sacks HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983; 309: 1358-61.
68. Chalmers TC, Matta RJ, Smith H Jr, Kunzler A-M. Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. *N Engl J Med* 1977; 297:1091-6.
69. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I. Medical. *Stat Med* 1989; 8: 441-54.
70. Miller JN, Colditz GA, Mosteller F. How study design affects outcomes in comparisons of therapy. II. Surgical. *Stat Med* 1989; 8: 455-66.
71. Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med* 1982; 72: 233-40.

## 8. Anhang

### 8.1 Auswertungsbogen

Kuss/Legler/Börgermann: page 1 from 5  
 Comparing Off-Pump and On-Pump CABG in Cardiac Surgery: A Meta-Propensity Score Analysis

---

**GENERAL INFORMATION**

Date of data extraction:                      Day      Month      Year  
   [ ][ ]    [ ][ ]    [ ][ ]

Reviewer: \_\_\_\_\_

**STUDY ELIGIBILITY ASSESSMENT**

Title: \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

Authors: \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

Source: \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

**STUDY CHARACTERISTICS**

	yes	unclear	no
1. Is an Off-Pump-group compared to an On-Pump-Group	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
2. Is the study a randomized trial?	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3. Is one of the clinical outcomes* given?	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
4. Are primary data used**?	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

\*death, stroke, myocardial infarction, atrial fibrillation, red blood cell transfusion, inotrope requirement, wound infection, re-operation for bleeding, acute renal failure, IABP support  
 \*\* e.g. a systematic review or a meta-analysis would use no primary data!

Include study       Exclude study

Notes: \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_







**STUDY RESULTS**

**Binary Outcomes**

Outcome	On-Pump		Off-Pump		Treatment effect (for Off-Pump)					
	Events	Number of Obs.	Events	Number of Obs.	Estimate	Value	SE	Lower 95%-CI	Upper	p
1 Short-term mortality										
2 Stroke										
3 Myocardial infarction										
4 Atrial fibrillation										
5 Red blood cell transfusion										
6 Inotrope Requirement										
7 Wound infection										
8 Re-operation for bleeding										
9 Acute renal failure										
10 IABP support										

## 8.2 Eingeschlossene Studien

Tabelle 4: Übersicht der eingeschlossene Studien (n=51)

Erstautor	Journal	Ch.	Ra.	Re.	Se.	Wi.	10/2006
<b>Al-Ruzzeah S</b>	BMJ 2006 Jun 10;332(7554):1365						<b>x</b>
<b>Alwan K</b>	Ann Thorac Surg 2004;77:2051-5.				<b>x</b>	<b>x</b>	
<b>Angelini GD</b>	Lancet 2002;359:1194 – 1199 ( <b>BHACAS I</b> )	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>	
<b>Angelini GD</b>	Lancet 2002;359:1194 – 1199 ( <b>BHACAS II</b> )	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>	
<b>Ascione R</b>	Circulation 2005;112:3833- 3838						<b>x</b>
<b>Ascione R</b>	Ann Thorac Surg 2006;81:97-103						<b>x</b>
<b>Baker</b>	Heart Surg Forum 2001;4(Suppl 1):S19-S23.	<b>x</b>	<b>x</b>		<b>x</b>	<b>x</b>	
<b>Carrier</b>	Heart Surg Forum 2003;6:E89-E92.	<b>x</b>	<b>x</b>		<b>x</b>	<b>x</b>	
<b>Cavalca V</b>	Ann Thorac Surg 2006;81:562-7						<b>x</b>
<b>Celik JB</b>	Ren Fail 2005; 27(2): 183-8						<b>x</b>
<b>Covino</b>	J Cardiovasc Surg (Torino) 2001;42:23-26	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>	
<b>Czerny M</b>	Ann Thorac Surg 2001;71:165-169.	<b>x</b>	<b>x</b>		<b>x</b>	<b>x</b>	
<b>Diegeler A</b>	Ann Thorac Surg 2000;69:1162- 1166.	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>	
<b>Gasz</b>	Exp Clin Cardiol 2004;9:26-30				<b>x</b>		
<b>Gasz B</b>	Eur Surg Res 2005;37:281-289						<b>x</b>
<b>Gerola</b>	Ann Thorac Surg 2004;77:569 –573.	<b>x</b>	<b>x</b>		<b>x</b>	<b>x</b>	
<b>Guler M</b>	Ann Thorac Surg 2001;71:152-157	<b>x</b>	<b>x</b>		<b>x</b>	<b>x</b>	
<b>Gulielmos V</b>	Eur J Cardiothorac Surg 2000;18:594- 601.		<b>x</b>		<b>x</b>		
<b>Johannson- Synnergren</b>	Scand Cardiovasc J 2004;38:53- 8.	<b>x</b>	<b>x</b>			<b>x</b>	
<b>Khan NE</b>	N Engl J Med 2004;350:21-28	<b>x</b>	<b>x</b>		<b>x</b>	<b>x</b>	
<b>Kobayashi</b>	Circulation 2005; 112(9 Suppl): 1338-1343				<b>x</b>		
<b>Kochamba GS</b>	Ann Thorac Surg 2000;69:1466- 1470.		<b>x</b>			<b>x</b>	
<b>Lee</b>	Ann Thorac Surg 2003;76:18 –26	<b>x</b>	<b>x</b>		<b>x</b>	<b>x</b>	
<b>Legare</b>	Circulation 2004;109:887- 892.	<b>x</b>	<b>x</b>		<b>x</b>	<b>x</b>	
<b>Lingaas</b>	Heart Surg Forum 2004;7:37-41.	<b>x</b>	<b>x</b>		<b>x</b>	<b>x</b>	
<b>Mariscalco G</b>	J Thorac Cardiovasc Surg 2006;131:1364-72						<b>x</b>
<b>Matata</b>	Ann Thorac Surg 2000;69:785-791.	<b>x</b>	<b>x</b>		<b>x</b>	<b>x</b>	
<b>Michaux I</b>	J Thorac Cardiovasc Surg 2006;131:1281-8						<b>x</b>



## 8.3 Ausgeschlossen Studien

Tabelle 5: Übersicht der ausgeschlossenen Studien und der Ausschlussgründe (n=56)

Erstautor	Journal	Ch.	Ra.	Re.	Se.	Wi	10/06	AG
van Dijk D	Circulation 2001;104:1761–1766.	x	x	X	x	x		DP
Ascione R	J Thorac Cardiovasc Surg 2001;121:689-696		x		x			DP
Puskas JD	J Thorac Cardiovasc Surg 2003;125:797–808.	x	x		x	x		DP
Ascione R	Ann Thorac Surg 1999;68:2237-2242.		x		x			DP
Lingaas	Ann Thorac Surg 2006;81:2089 –96						x	DP+NO
Ostergaard Jensen B	Circulation 2006;113;2790-2795						x	DP
Ascione R	Circulation 2000;102:1530–1535.		x	X	x			DP
Al-Ruzzeh S	Heart Surg Forum 2003;6:89-93		x		x			NO
Anderson RE	Eur J Cardiothorac Surg 28 (2005) 425–430						x	NO
Arom KV	Ann Thorac Surg. 2000;69:704-10.		x					NR
Ascione R	Ann Thorac Surg 1999;68:493–498		x	X	x			DP
Ascione R	Eur J Cardiothorac Surg 1999;15:685-690.		x		x			DP
Ascione R	Ann Thorac Surg 2000;69:1198-1204.		x		x			DP
Ascione R	Eur Heart J 2004;25:765-770.		x		x			DP
Aydin NB	J Thorac Cardiovasc Surg 2003;125:843-848.		x					NO+NR
Bull DA	Ann Thorac Surg 2001; 71:170-175		x					NR
Caputo	Ann ThoracSurg 2002;74:2088 –96.	x		X	x	x		DP
Cimen S	Eur J Cardiothorac Surg 2003;23:589-594.		x					NR
Cox CM	Ann Thorac Surg 2000;69:140-145		x		x			DP
Czerny M	Eur J Cardiothorac Surg 2000;17:737-742	x	x		x			DP
Diegeler A	Eur J Cardiothorac Surg 1999; 15:680-684					x		DP+NR+OT
Diegeler A	Circulation 2000;102(19 Suppl 3):III95-III100.		x					DP+NR+OT
Dorman	J Cardiothorac Vasc Anesth 2004;18:25-29	x	x		x			NO
Ernest CS	Ann Thorac Surg 2006;81:2105–14						x	NO
Gerritsen	Eur J Cardiothorac Surg 2001; 20:923-929		x					NR
Gu YJ	Ann Thorac Surg 1998;65:420–424	x	x		x	x		OT
Gulielmos V	Eur J Cardiothorac Surg 1999;16:S34 –	x	x		x	x		DP



## 9. Thesen

1. Der heutige Qualitätsstandard zur Evaluation von Therapieeffekten sind Metaanalysen randomisierter, kontrollierter, klinischer Studien. Es gibt jedoch Situationen, in denen die Durchführung randomisierter, kontrollierter, klinischer Studien unnötig, ungeeignet, unmöglich, unzureichend oder unethisch ist.
2. Eine Alternative stellen die nicht randomisierten Beobachtungsstudien dar, bei denen Confounding jedoch ein entscheidendes Problem ist.
3. Die Propensity-score-Analyse ist eine sehr valide Methode zur Confounderadjustierung bei der Datenauswertung nicht randomisierter Studien. Dabei gibt der PS die Wahrscheinlichkeit an, als Studienproband der Therapiegruppe anzugehören bzw. die zu untersuchende Intervention zu erhalten.
4. Es sind bisher eine Vielzahl von Untersuchungen durchgeführt worden, die klären sollten, ob sich die aus randomisierten Studien hervorgegangenen Therapieeffekte von denen aus nicht randomisierten Studien unterscheiden. Es konnte bisher nicht zufrieden stellend bewiesen werden ob beide Studienformen gleiche Ergebnisse produzieren bzw. die Ergebnisse durch eine Studienform über- oder unterschätzt werden.
5. Im Rahmen der vorliegenden Arbeit werden randomisierte Studien mit Propensity-score-Analysen in Bezug auf Ihre Ergebnisse miteinander verglichen. Das gemeinsame Thema der zu vergleichenden Studien ist das wohl derzeit am meisten debattierte und am stärksten polarisierende im Bereich der Kardiochirurgie, On-Pump versus Off-Pump
6. Bei der vorliegenden Arbeit handelt es sich um eine Meta-Propensity-score-Analyse. Da Confounding auch bei dem Vergleich von Studien eine Rolle spielt, wurde die PS-Methode zur Meta-Confounderadjustierung genutzt.
7. Es wurden die Literaturverzeichnisse der fünf größten Übersichtsarbeiten zum Thema On-Pump versus Off-Pump sowie die Datenbank von MEDLINE verwendet, um die randomisierten, kontrollierten klinischen Studien für den Vergleich zusammenzutragen. Die Propensity-score-Analysen wurden einer Meta-Analyse entnommen.
8. Die Datenextraktion erfolgte durch zwei unabhängige Untersucher mittels eines eigens für diesen Zweck entwickelten Auswertungsbogens. Neben Studien- und Patienteneigenschaften wurden zehn für den Vergleich essentiellen binären Outcomes extrahiert.

9. Für den Vergleich der Behandlungseffekte beider Therapiemethoden und damit der Studienformen wurden die Differenzen der logarithmischen Odds Ratios und die dazugehörigen 95%-Konfidenzintervalle ermittelt.
10. In den Vergleich gingen die Daten aus 51 randomisierten Studien und 28 Propensity-Score-Analysen ein. Nach dem Meta-Matching konnten 10 PS-Analysen (25.552 Patienten) und 29 RCTs (2.723 Patientin) valide gegenübergestellt werden, da der Unterschied zwischen den Studien- und Patienteneigenschaften minimiert bzw. aufgehoben wurden.
11. Die Meta-Odds-Ratios zeigen sowohl in den 29 RCTs als auch 10 PS-Analysen einen signifikant besseren Therapieeffekt auf Seiten der Off-Pump-Gruppe.
12. Die Differenzen der Meta-Odds-Ratios beider Studientypen in der gematchten Gruppe zeigen sowohl in der Einzeldarstellung als auch in der Gesamtheit keinen signifikanten Unterschied in den Ergebnissen aus randomisierten Studien und PS-Analysen.
13. Die Propensity-score-Analyse produziert im vorliegenden Fall identische Ergebnisse wie eine randomisierte, kontrollierte, klinische Studie und ist somit eine ebenbürtige Methode zur Evaluation von Therapieeffekten.
14. Zur Stützung der Ergebnisse dieser Arbeit sollte ein Vielzahl weiterer derartiger Vergleiche durchgeführt, vor allem aber auch am Beispiel pharmakologischer Behandlungsformen.

# Lebenslauf

---

Thomas Legler,  
\*12.03.1981 in Pirna  
ledig

## **Berufserfahrung:**

---

08/2005 – 09/2005	Famulatur in der Viszeralchirurgie, Marta-Maria-Krankenhauses Halle/Dörlau
07/2006 – 08/2006	Famulatur in der Kardiochirurgie, Herzzentrum Dresden
08/2006 – 09/2006	Famulatur in der Traumatologie/Orthopädie, Klinikum Pirna GmbH
02/2007	Famulatur im ambulanten Bereich, Praxis für Kinder- und Jugendmedizin
<b>08/2007 – 08/2008</b>	<b>Praktisches Jahr:</b>
08/2007 – 12/2007	Orthopädisches Terial im Universitätsklinikum Halle
12/2007 – 04/2008	Internistisches Terial im Klinikum Pirna
04/2008 – 08/2008	Chirurgisches Terial im Klinikum Pirna
Seit 01.01.2009	Arzt in Weiterbildung Klinik für Chirurgie II Klinikum Pirna

---

## **Bildung:**

---

2008	Staatsexamen des Studiums der Humanmedizin
Seit 2006	Promotion: „Gibt es einen Unterschied zwischen randomisierten und nicht-randomisierten Studien? – Evidenz einer „Meta-Propensity-score-Analyse“ in der Herzchirurgie“; Institut für Medizinische Epidemiologie und Biometrie der MLU Halle/Wittenberg; Betreuer: Dr. O. Kuss
2004	Ärztliche Vorprüfung
2002 - 2008	Studium der Humanmedizin Martin-Luther-Universität Halle/ Wittenberg
03/2000 – 09/2001	Ausbildung zum Physiotherapeuten Berufsschule des Universitätsklinikums „Carl-Gustav-Carus“ der TU Dresden
1995 – 1999	Rainer – Fetscher – Gymnasium Pirna, Abitur
1992 – 1995	Johann – Gottfried – Herder – Gymnasium Pirna
1989 – 1992	Polytechnische Oberschule „Alfred Vater“ Pirna
1987 – 1989	Polytechnische Oberschule „Georg Haak“ Pirna

---

## **Nebentätigkeiten:**

---

1999 - 2002	Mitarbeit in der mütterlichen physiotherapeutischen Praxis Arbeit in der Systemgastronomie
-------------	---

---

## **Selbständigkeitserklärung und Erklärung über frühere Promotionsversuche**

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässig Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten sind unter Angabe der Quelle gekennzeichnet.

Ich versichere, dass ich für die inhaltliche Erstellung der vorliegenden Arbeit nicht die entgeltliche Hilfe von Vermittlungs- und Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen habe. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Ich erkläre, dass ich keinerlei frühere Promotionsversuche unternommen habe und dass an keiner anderen Fakultät oder Universität ein Promotionsverfahren anhängig ist.

.....  
Thomas Legler