

Aus dem Institut für Medizinische Epidemiologie, Biometrie und Informatik der  
Medizinischen Fakultät der Martin-Luther-Universität Halle-Wittenberg

(Direktor: Prof. Dr. rer. nat. habil. Johannes Haerting)

Die externe Validität „gematchter“ Propensity-Score-Analysen -  
Ein systematisches Review anhand eines Beispiels aus der  
Herzchirurgie

Dissertation zur Erlangung des akademischen Grades Doktor der Medizin (Dr. med.)

vorgelegt der Medizinischen Fakultät der Martin-Luther-Universität Halle-Wittenberg

von Lisa Angelika Redies  
geboren am 03.07.1989 in Boston (USA)

Betreuer

Prof. Dr. sc. hum. Oliver Kuß

Gutachter

1. Prof. Dr. sc. hum. Oliver Kuß
2. Prof. Dr. med. Jan Gummert (Bad Oeynhausen)
3. Prof. Dr. Markus Neuhäuser (Koblenz)

Datum der Eröffnung des Promotionsverfahrens

11.02.2013

Datum der Verteidigung der Dissertation

27.01.2014

## Referat

Bei der Prüfung von Therapien in vergleichenden Studien kann auf verschiedene Verfahren zur Bildung von Behandlungsgruppen zurückgegriffen werden, um eine möglichst hohe interne und externe Validität der späteren Ergebnisse sicherzustellen. In dieser Arbeit wurde untersucht, wie und ob Randomisierung bzw. Propensity-Score-Matching Studienkollektive einschränken und ob sich daraus grundsätzliche Aussagen zur externen Validität von Studien, die diese Verfahren anwenden, ableiten lassen. Das vorliegende systematische Review wurde mit Studien durchgeführt, die ihrerseits verschiedene Zielgrößen bei On- bzw. Off-Pump-Bypass-Operationen untersuchten. Es wurden für die Merkmale Alter, linksventrikuläre Ejektionsfraktion (EF), Geschlecht, Diabetes, Hypertonie und Myokardinfarktgeschichte (MI) gemittelte Lagemaße für randomisiert-kontrollierte Studien (RCT) und für die „gematchte“ Population der Propensity-Score-Studien (PSM) ermittelt und diese mit der Grundgesamtheit der Patienten verglichen. Des Weiteren wurde untersucht, inwieweit sich Kollektive durch Propensity-Score-Matching verändern.

Es stellte sich heraus, dass die externe Validität von RCTs bezüglich der Ejektionsfraktion eingeschränkt ist. Bezüglich der anderen 5 Merkmale konnten nur vernachlässigbar kleine bzw. klinisch nicht relevante Differenzen zur Grundgesamtheit festgestellt werden. Die externe Validität „gematchter“ Propensity-Score-Studien zeigte sich in keinem der sechs geprüften Merkmale klinisch relevant eingeschränkt. Es werden durch Matching insbesondere Männer, junge Patienten und solche mit niedriger Ejektionsfraktion nicht in das Studienkollektiv mit aufgenommen.

Die „gematchten“ Propensity-Score-Studien sind im vorliegenden systematischen Review den RCTs bezüglich ihrer externen Validität überlegen. Die externe Validität von „gematchten“ Propensity-Score-Studien zeigt sich nicht eingeschränkt. Nachteilig jedoch bleibt, dass die *interne* Validität „gematchter“ Propensity-Score-Studien im Gegensatz zu RCTs auf Grund von nicht zu garantierender Strukturgleichheit unsicher ist.

Redies, Lisa Angelika: Die externe Validität „gematchter“ Propensity-Score-Analysen - Ein systematisches Review anhand eines Beispiels aus der Herzchirurgie, Halle (Saale), Martin-Luther-Universität Halle-Wittenberg, Medizinische Fakultät, Institut für Medizinische Epidemiologie, Biometrie und Informatik, 77 Seiten, 2013.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Material und Methoden</b>	<b>5</b>
2.1	Grundlagen . . . . .	5
2.1.1	Die koronare Herzkrankheit und ihre Therapie . . . . .	5
2.1.2	Der Propensity-Score als Werkzeug zur Erstellung von Behandlungsgruppen . . . . .	7
2.2	Auswahl der Daten . . . . .	8
2.2.1	Auswahl der RCT-Studien . . . . .	8
2.2.2	Auswahl der PS-Studien . . . . .	9
2.2.3	Extrahierte Daten . . . . .	9
2.3	Hauptanalyse . . . . .	9
2.3.1	Einheitliche Lagemaße . . . . .	10
2.3.2	Fehlende, imputierte und vollständige Daten . . . . .	12
2.3.3	Externe Validität - Between-Vergleich . . . . .	12
2.3.4	Externe Validität - Within-Vergleich . . . . .	14
2.4	Sensitivitätsanalyse . . . . .	14
2.4.1	Einheitliche Lage- und Streuungsmaße . . . . .	15
2.4.2	Externe Validität - Between-Vergleich . . . . .	16
<b>3</b>	<b>Ergebnisse</b>	<b>17</b>
3.1	Ausgewählte Studien . . . . .	17
3.1.1	RCT-Studien . . . . .	17
3.1.2	PS-Studien . . . . .	17
3.2	Extrahierte Daten . . . . .	18
3.2.1	Umgang mit unplausiblen und problematischen Daten . . . . .	20
3.3	Hauptanalyse . . . . .	20
3.3.1	Externe Validität - Between-Vergleich . . . . .	20
3.3.2	Externe Validität - Within-Vergleich . . . . .	25
3.4	Sensitivitätsanalyse . . . . .	33
<b>4</b>	<b>Diskussion</b>	<b>36</b>
<b>5</b>	<b>Zusammenfassung</b>	<b>43</b>
<b>6</b>	<b>Literaturverzeichnis</b>	<b>45</b>

<b>7</b>	<b>Anhang</b>	<b>48</b>
7.1	Eingeschlossene Studien . . . . .	48
7.1.1	Randomisiert-kontrollierte Studien . . . . .	48
7.1.2	Propensity-Score-Studien . . . . .	57
7.2	Tabellen und Grafiken . . . . .	61
<b>8</b>	<b>Thesen</b>	<b>68</b>
<b>9</b>	<b>Tabellarischer Lebenslauf</b>	
<b>10</b>	<b>Selbstständigkeitserklärung</b>	
<b>11</b>	<b>Erklärung über frühere Promotionsversuche</b>	
<b>12</b>	<b>Danksagung</b>	

## Abkürzungsverzeichnis

<b>EF</b>	Linksventrikuläre Ejektionsfraktion
<b>General</b>	Grundgesamtheiten der PS-Studien
<b>GLM</b>	Generalized linear model
<b>GLMM</b>	Generalized linear mixed model
<b>MA</b>	Metaanalyse
<b>MI</b>	Myokardinfarktgeschichte
<b>PS</b>	Propensity-Score
<b>[PSxx]</b>	verweist auf die im Anhang aufgeführten eingeschlossenen Propensity-Score-Studien
<b>PSM</b>	„gematchte“ Population der PS-Studien
<b>RCT</b>	randomisiert-kontrollierte Studien (engl.: randomized controlled trial)
<b>[RCTxx]</b>	verweist auf die im Anhang aufgeführten eingeschlossenen randomisiert-kontrollierten Studien

# 1 Einleitung

Bei der Erstellung klinischer Studien, die den Einfluss verschiedener Therapien auf eine Zielgröße untersuchen, werden Behandlungsgruppen gebildet. Diese werden unterschiedlichen Therapien zugeführt und die Ausprägung der Zielgrößen anschließend verglichen. Damit man den Effekt der Therapien auf die Zielgröße ohne störende Einflüsse (sog. Confounder) untersuchen und auf die durchschnittliche Klinikpopulation übertragen kann, sollten die Gruppen so gestaltet sein, dass die Studie valide Ergebnisse erzeugt. Es werden interne und externe Validität unterschieden. Eine Studie gilt dann als intern valide, wenn die gewonnenen Ergebnisse und die daraus abgeleiteten Schlussfolgerungen für die Studienpopulation korrekt sind. Extern valide ist eine Studie dann, wenn die Ergebnisse auch gültig für die entsprechende Klinikpopulation (Gesamtheit *aller* Patienten mit dieser Erkrankung) sind, die Schlussfolgerungen also korrekt generalisiert werden können. Bei der Übertragung von Studienergebnissen auf die Klinikpopulation ist es von großer Wichtigkeit zu wissen, wie es um die externe Validität einer Studie bestellt ist und für welches Kollektiv die Studie Aussagen machen kann.

## Randomisiert-kontrollierte Studien

Randomisierung gilt derzeit als Goldstandard zur Prüfung von Therapien. Es werden die Studienteilnehmer zufällig auf zwei oder mehrere Gruppen verteilt. Somit wird nicht nur für bekannte, sondern auch für evtl. unbekannt, aber einflussreiche Confounder balanciert. Die Strukturgleichheit der Gruppen sichert die interne Validität. Bezüglich der externen Validität randomisiert-kontrollierter Studien wird jedoch seit längerem kritisiert, dass sie oft eine rigorose Auswahl des Patientenkollektivs vornehmen. Dabei werden insbesondere häufig Patienten auf Grund von Alter oder Grunderkrankungen ausgeschlossen. Als medizinische Gründe werden z. B. das hohe Risiko von Nebenwirkungen für bestimmte Patientengruppen angegeben [1]. Wissenschaftlich betrachtet reduziert man das Risiko des Bias, indem man von Anfang an Patienten ausschließt, die höchst wahrscheinlich die Studie abbrechen werden; man reduziert den sogenannten „loss to follow up“ [1]. Dieser systematische Ausschluss birgt erhebliche Risiken in der Anwendung der gewonnenen Studienergebnisse: Wenn nicht eindeutig hervorgehoben ist, für welche Patienten die Ergebnisse gültig sind, werden eventuell falsche Grundsätze generalisiert und Patienten können Schaden durch eine für sie unpassende Behandlung nehmen. Häufig bringen Ärzte im klinischen Alltag den Ergebnissen von RCTs ein solches Misstrauen bezüglich der Generalisierbarkeit entgegen, dass sie aus Scheu die möglicherweise durchaus angemessene neuere Therapieform gerade *nicht* anwenden [2]. Des Weiteren werden durch den systematischen Ausschluss mancher Gruppen, wie z.B. hoch komorbider Patienten, Schwangerer und Kindern, keine Erkenntnisse für diese Gruppen von vergleichsweise schwierigen Patienten generiert. Auf Grund unzureichender wissenschaftlicher Belege sind die Therapiemöglichkeiten für sie oft eingeschränkt. Es gibt zahlreiche Studien, die eine Diskrepanz zwischen normaler

Klinikpopulation und denjenigen Probanden feststellen, die in RCTs eingeschlossen werden [3, 4]. In einer Studie von Masoudi et al. [5] wird sogar festgestellt, dass von den dort untersuchten 20 388 Patienten einer Klinikpopulation nur 18 – 25% die Aufnahmekriterien für drei grundlegende RCTs über Stauungsinsuffizienz des Herzens erfüllten. Es ist außerdem bekannt, dass häufig ausgeschlossen wird, ohne die genauen Ausschlusskriterien ausreichend zu dokumentieren [6]. Dies kann zu den bereits oben genannten Fehlübertragungen auf falsche Patientenkollektive oder Misstrauen gegenüber RCTs führen. Allerdings muss an dieser Stelle auch hervorgehoben werden, dass durchaus nicht alle RCTs einen Mangel an externer Validität aufweisen. So stellten Sorg et al. [7] in einer Metaanalyse (MA) von RCTs zu adjuvanter Therapie bei Pankreas-Adenokarzinom fest, dass trotz kleinerer Einschränkungen die externe Validität der untersuchten Studien als durchaus akzeptabel zu bewerten ist.

In einer Metaanalyse zur externen Validität von randomisierten Studien wurden weiterhin folgende Beobachtungen angestellt [1]: Die Generalisierbarkeit von randomisierten Studien ist eingeschränkt, da die in die Studien eingeschlossenen Patienten im Durchschnitt eine *schlechtere* Prognose haben, als die ausgeschlossenen. Probanden, die an randomisiert-kontrollierten Studien zur Bewertung von Behandlungsmöglichkeiten teilnehmen, sind weniger wohlhabend, gebildet und gesund als solche, die *nicht* teilnehmen. Das Gegenteil lässt sich feststellen bei Studien zur Bewertung von Präventivmaßnahmen: hier sind die Teilnehmer gesünder, wohlhabender und gebildeter als die Nicht-Teilnehmer. Auf der einen Seite wird vermutet, dass auf Grund der Tatsache, dass in randomisiert-kontrollierten Behandlungsstudien eher Probanden teilnehmen, die sich in einem schlechteren Gesundheitszustand befinden und so deutlicher von einer Behandlung profitieren, die Effekte der Behandlung tendenziell *überschätzt* werden. Auf der anderen Seite wird auf Grund der überdurchschnittlich guten Verfassung der Teilnehmer an Präventivstudien der Effekt der Präventivmaßnahmen hier *unterschätzt* [1].

## **Nicht-randomisierende Verfahren zur Balancierung von Behandlungsgruppen**

Alternativ kann mit nicht-randomisierenden Methoden zur Erstellung von Behandlungsgruppen gearbeitet werden. Um Confounding zu vermeiden, kann zum Beispiel nach Beendigung der ersten statistischen Auswertung eine sog. Regressions-Analyse durchgeführt werden. Mittels linearer Regression für metrische Merkmale bzw. logistischer Regression für binäre Merkmale werden die *direkten* Einflüsse der Variablen auf die Zielgröße errechnet, indem die wechselseitigen Einflüsse der Variablen untereinander bereinigt werden. Die Regressions-Analyse lässt sich jedoch schwer anwenden, wenn die Zielgröße von Interesse selten ist [8]. Des Weiteren wird nur der Einfluss derjenigen Zielgrößen bereinigt, die dem Studienleiter bekannt sind. Für möglicherweise nicht beachtete Confounder wird nicht adjustiert.

1983 publizierten Rosenbaum und Rubin eine neue Methode zur Prüfung von Therapien: den Propensity-Score (PS) [9]. Der Propensity-Score findet immer häufiger Anwendung in nicht-

randomisierten oder -randomisierbaren medizinischen Studien, um Confounding und systematische Unterschiede zwischen Behandlungsgruppen (sog. Selektionsbias) zu minimieren und so die interne Validität der Studien zu verbessern. Der Propensity-Score gibt definitionsgemäß die Wahrscheinlichkeit für einen Studienteilnehmer an, Therapie A und nicht Therapie B zu erhalten. Es werden Merkmale, die der Studienleiter für relevant für die Therapieentscheidung hält, bzw. deren Ausprägungen, zu einer einzigen Zahl - dem Propensity-Score - zusammengefasst. Von jedem Studienteilnehmer liegt eine Fülle von Daten vor; es wird jedoch für jedes übergeordnete Feld nur ein Merkmal als Vertreter ausgewählt (z.B. die Körpergröße als Vertreter für BMI, Gewicht, Körperoberfläche etc.). Für die ausgesuchten Merkmale wird mit Hilfe logistischer Regression ein Koeffizient ermittelt. Die Summe aus dem Produkt der Koeffizienten mit den einzelnen Werten ergibt nach Anwendung der Expit-Transformation den PS [8, 10]. Wenn über viele Probanden gemittelt wird, ähneln sich Probanden mit ähnlichen Propensity-Scores in den zuvor festgesetzten Merkmalen.

Mit Hilfe des Propensity-Scores können unter Verwendung verschiedener Verfahren Behandlungsgruppen erstellt werden. Eine ausführlichere Beschreibung dieser Verfahren - Regressionsanalyse, Stratifizierung, Matching und Gewichtung mit inverser Behandlungswahrscheinlichkeit - findet sich in Kapitel 2.1.2. In der vorliegenden Arbeit beschränke ich mich auf die Betrachtung des Matchings mit Hilfe des Propensity-Scores (im Folgenden werden die Termini „Matchen“ bzw. „gematcht“ als eingedeutscht verstanden und auf Anführungszeichen verzichtet). Dabei verteilt man Paare möglichst ähnlichen PSs auf zwei verschiedene Behandlungsgruppen und erhält so zwei Gruppen, die sich in der Ausprägung der untersuchten Merkmale ähnlich sind. Beim Matchen anhand des PS ist es möglich, aus einem bestehenden großen Patientenkollektiv retrospektiv den PS zu ermitteln, gleichmäßige Gruppen zu bilden und die Einflüsse verschiedener Therapien bzw. Expositionen auf theoretisch immer neu wählbare Zielgrößen zu untersuchen.

Nicht-randomisierten Studien im Allgemeinen wird ein Nachteil in der internen Validität nachgesagt. Dadurch, dass in allen genannten nicht-randomisierenden Verfahren stets nur für *bekannte* Confounder balanciert wird, ist nicht sicher auszuschließen, dass es weitere Confounder gibt, die den wahren Effekt der Therapie auf die Zielgröße verschleiern und so die interne Validität verschlechtern. Zudem ist zumindest für gematchte Propensity-Score-Studien bekannt, dass oft nur mangelhaft auf erfolgreiches Matchen und eine zufriedenstellende Balancierung der Behandlungsgruppen hin geprüft wird. So wird in einer Arbeit von Austin [11] festgestellt, dass in keiner einzigen der dort untersuchten gematchten PS-Studien ein adäquates Rechenmodell zur Überprüfung der Balancierung der Behandlungsgruppen verwendet wurde. Des Weiteren gaben 51% der gesichteten Studien gar keine Informationen dazu an, unter welchen Prämissen und Annahmen gematcht wurde. Die interne Validität kann durch eine unzureichende Balancierung empfindlich eingeschränkt werden.

Als Vorteil nicht-randomisierter Studien wird angesehen, dass hier *nicht* selektiv bei der Aus-

wahl der Probanden vorgegangen wird. In der großen Mehrzahl der nicht-randomisierenden Verfahren zur Prüfung von Therapien werden weder auf Grund taktischer Überlegungen (siehe Abschnitt *Randomisiert-kontrollierte Studien*, [1]) noch strukturell durch die Methode bedingt Probanden von der Studie ausgeschlossen. Es ist davon auszugehen, dass die durchschnittlichen Eigenschaften der Studienpopulation hier denen der Klinikpopulation entsprechen. Die externe Validität ist nicht so systematisch eingeschränkt wie bei RCTs. Eine Ausnahme bildet das Propensity-Score-Matching. Hier fallen nicht matchbare Patienten - meist solche mit besonders hohem oder niedrigem PS - heraus und werden nicht in das Studienkollektiv aufgenommen. Ob und wie die durchschnittlichen Eigenschaften des Kollektivs verändert werden (und evtl. die externe Validität negativ beeinflussen) wurde bisher noch nicht untersucht. Bezüglich der externen Validität von nicht-randomisierten Studien im Allgemeinen (im Vergleich zu RCTs) stellten McKee et al. [1] des Weiteren fest, dass insbesondere in Studien zu Präventiv-Maßnahmen eine überdurchschnittlich hohe Zahl von Probanden eingeschlossen wurde, die auf Grund ihres Gesundheitsverhaltens eine höhere Chance hatten, von den angegebenen Maßnahmen zu profitieren. Während RCTs vor allem in universitären Zentren durchgeführt werden, schließen nicht-randomisierte Studien häufiger nicht-universitäre Zentren ein [1]. Es ist anzunehmen, dass sich die Patienten in universitären Zentren auf Grund unterschiedlicher technischer Standards und medizinischer Möglichkeiten von denen nicht-universitärer Zentren unterscheiden. Die externe Validität sowohl von RCTs als auch von nicht-randomisierten Studien könnte darunter leiden.

### **Zentrale Fragestellung**

Es stellt sich die Frage, ob die neuere Methode des Matchens anhand des PS die zentralen Kritikpunkte der Randomisierung bezüglich der externen Validität umgehen kann. Die in der vorliegenden Arbeit zu beantwortende Frage lautet: Ist das Matchen anhand des Propensity-Scores eine nicht nur einfachere (da auch retrospektiv durchzuführende) und kostengünstigere Methode zur Erstellung von Behandlungsgruppen, sondern vielleicht auch eine, die eine höhere externe Validität aufweist als der derzeitige Goldstandard der Randomisierung? Lassen sich mit diese Methode also eventuell Ergebnisse produzieren, die einer größeren Anzahl von Patienten als bisher zu Gute kommen können?

Im Folgenden soll untersucht werden, ob sich gewisse Merkmale in randomisiert-kontrollierten Studien und in gematchten Propensity-Score-Studien von der Grundgesamtheit grundsätzlich unterscheiden. Zudem soll geprüft werden, ob durch Matchen in Propensity-Score-Studien die Merkmale der zu Grunde liegenden Patientenpopulation grundsätzlich verändert werden. Diese Fragen zur externen Validität der Studientypen werden im Folgenden anhand eines systematischen Reviews von 91 randomisiert-kontrollierten und 38 PS-Studien, die sich mit dem Vergleich von On- vs. Off-Pump Bypass-Operationen beschäftigen, überprüft.

## 2 Material und Methoden

### 2.1 Grundlagen

#### 2.1.1 Die koronare Herzkrankheit und ihre Therapie

Die für diese Arbeit herangezogenen Studien untersuchten verschiedene Zielgrößen bei Bypass-Operationen mit On-Pump bzw. Off-Pump-Verfahren. Im folgenden Abschnitt soll die koronare Herzkrankheit als wichtige Indikation der koronaren Bypass-Operationen, sowie die Grundlagen dieser Operationstechniken erklärt werden.

**Die koronare Herzkrankheit** Wie in den meisten Industrienationen, sind in Deutschland Herz-Kreislauf-Erkrankungen, und unter diesen insbesondere die koronare Herzkrankheit, die häufigste Todesursache unter Erwachsenen. Die koronare Herzkrankheit bezeichnet eine Atherosklerose der Herzkranzarterien und die daraus resultierende Sauerstoffminderversorgung des distal der Stenose gelegenen Myokards. Diese Ischämie des Gewebes kann in verschiedenen Formen in Erscheinung treten, wie z.B. der Angina pectoris, dem Myokardinfarkt oder dem plötzlichen Herztod. Als Risikofaktoren gelten Zigarettenrauchen, arterielle Hypertonie, LDL-Cholesterin-Erhöhung, HLD-Cholesterin-Erniedrigung, Lebensalter (Männer  $\geq 45$  J., Frauen  $\geq 55$  J.), positive Familienanamnese, Diabetes mellitus sowie körperliche Inaktivität. In Deutschland gaben 2010 7% der Frauen und knapp 10% der Männer an, dass bei ihnen eine koronare Herzkrankheit diagnostiziert wurde [12, S. 96-98].

Therapeutisch wird der koronaren Herzkrankheit auf verschiedenen Stufen begegnet. Die Prävention zielt auf Ausschaltung der Risikofaktoren (Nikotinabstinenz, körperliche Aktivität etc.) bzw. auf eine gute Einstellung der Grunderkrankungen, wie der Hypertonie oder des Diabetes mellitus. Die im akuten Koronarsyndrom zusammengefasste instabile Angina pectoris, der ST-Strecken-Hebungsinfarkt (STEMI) und der nicht-ST-Strecken-Hebungsinfarkt (NSTEMI) bedürfen einer intensivmedizinischen Versorgung und ggfs. einer Revaskularisationstherapie (siehe unten). Die stabile Angina pectoris kann medikamentös durch Thrombozytenaggregationshemmer, Betarezeptorenblocker, Nitrate oder Kalziumantagonisten beherrscht werden. Sollte sich die Angina-pectoris-Symptomatik unter Medikation nicht ausreichend bessern, kann ein Revaskularisationsversuch unternommen werden. Mittels perkutaner transluminaler koronarer Angioplastie werden die stenosierte Koronargefäße meist entweder mit einem Ballon erweitert oder es wird dauerhaft ein sogenannter Stent zur Revaskularisation eingesetzt. Als Indikationen gelten Erkrankungen von 1- bis 3-Gefäßen mit signifikanter Stenose ( $> 70\%$ ) sowie der akute oder chronische Verschluss. Liegt jedoch eine Hauptstammstenose der linken Koronararterie, eine Mehrgefäßerkrankung mit komplexer Stenose oder eine 2- bzw. 3- Gefäßerkrankung mit Beteiligung des proximalen Ramus interventricularis anterior vor, ist die operative Koronarrevaskularisation indiziert. Hierbei stehen verschiedene Techniken zur Verfügung. Zu den klassischen Methoden zählen die On- bzw. Off-Pump-Bypass-Operation. Eine

genauere Beschreibung der Operationstechnik findet sich im nächsten Absatz. In der vorliegenden Arbeit werden ausschließlich Studien untersucht, die sich mit Zielgrößen von Off- bzw. On-Pump-Bypass-Operationen befassen.

Die Prognose der koronaren Herzkrankheit ist abhängig vom Funktionszustand des linken Ventrikels, vom Ausmaß der Myokardischämie und der Lokalisation bzw. der Zahl der betroffenen Gefäße. Die Letalitätsrate pro Jahr ohne Revaskularisationstherapie beträgt bei 1-Gefäßerkrankung 3 – 4%, bei Hauptstammstenose der linken Herzkranzarterie > 30%. Sofern nicht anders gekennzeichnet, ist der Abschnitt „Die koronare Herzkrankheit“ angelehnt an Herold [13, S. 239-251].

**Technik der On- bzw. Off-Pump-Bypass-Operation** Die operative Therapie zielt auf eine Umgehung der stenosierten Abschnitte der Gefäße mittels Bypass. Verwendet werden Beinvenen oder die Arteria thoracica interna. In Abhängigkeit der Risikofaktoren (multimorbide Patienten, verkalkte Aorta etc.) entscheidet sich der Chirurg für ein On-Pump- oder Off-Pump-Verfahren. Die Anlegung eines sog. aortokoronaren Bypasses (OPS-Code 5-361) wurde in Deutschland im Jahr 2010 83 943 mal durchgeführt [14].

Beim On-Pump-Verfahren versorgt eine extrakorporale Zirkulation mittels Herz-Lungen-Maschine den Körper unter Umgehung des kleinen Kreislaufs mit oxygeniertem Blut. Dazu wird die Hohlvene kurz über und unter dem rechten Ventrikel unterbunden und das venöse Blut über die Herz-Lungen-Maschine in den Aortenbogen zurückgeleitet. Eine Blutpumpe ersetzt die Pumpfunktion des Herzens und ein Oxygenator reichert das Blut mit Sauerstoff an und entfernt zur gleichen Zeit das  $CO_2$ . Über einen Wärmeaustauscher wird die gewünschte Körpertemperatur aufrechterhalten; ein Sauger führt auch das intraoperativ vom Herzen abgesaugte Blut in die Herz-Lungen-Maschine und damit in den Körperkreislauf zurück (siehe schematische Darstellung der extrakorporalen Zirkulation in Abbildung 1 auf Seite 7). Das Herz wird mittels kardiopleger Lösung elektromechanisch entkoppelt und mittels Eiswasser gekühlt, sodass der Stoffwechsel angehalten wird, keine schädlichen Metabolite entstehen können und das Herz während der Operation nicht schlägt. Im Gegensatz dazu ist die Off-Pump-Operation eine Operation am schlagenden Herzen ohne Herz-Lungen-Maschine. Stabilisatoren - zumeist zwei Bügel - halten die Operationsstelle still, die operierten Koronargefäße werden proximal und distal gedrosselt (mittels Tourniquetnähren). Dieses Verfahren ist insbesondere für Patienten mit den oben genannten Risikofaktoren zu bevorzugen. Sofern nicht anders gekennzeichnet, ist der Abschnitt „Technik der On- bzw. Off-Pump-Bypass-Operation“ angelehnt an Henne-Bruns et al. [15, S. 1012 - 1017].

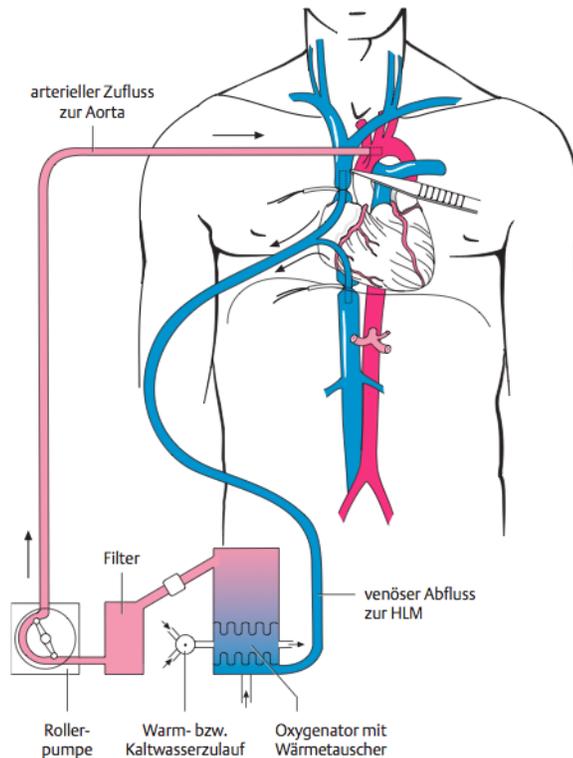


Abbildung 1: Schematische Darstellung der extrakorporalen Zirkulation [16, S. 949].

### 2.1.2 Der Propensity-Score als Werkzeug zur Erstellung von Behandlungsgruppen

Wie bereits beschrieben, gilt die Randomisierung derzeit als Goldstandard zur Prüfung von Therapien, da sie mit wachsender Probandenzahl eine annähernd gleiche Verteilung der Merkmale erreicht und gleichzeitig auch für nicht bekannte Confounder balanciert.

Der von Rosenbaum und Rubin publizierte Propensity-Score [9] lässt sich mittels verschiedener Verfahren zur Vermeidung von Confounding und Selektionsbias von Behandlungsgruppen anwenden. Mit dem errechneten PS kann man mittels folgender Methoden weiter verfahren (angelehnt an eine Arbeit von Austin [17]):

**Regressions-Analyse** Der Propensity-Score wird als einfache Kovariable in der Regressions-Analyse genutzt. Der Effekt der Therapie auf die Zielgröße wird für den PS bereinigt. Es lässt sich der Zusammenhang zwischen dem PS und der Zielgröße darstellen.

**Stratifizierung** Man bildet anhand des PS Quantilen und vergleicht Therapie A und Therapie B miteinander nur innerhalb dieser Quantilen, also in Gruppen von Probanden mit annähernd gleichen Merkmalen. Dies gestaltet sich mit dem Propensity-Score weitaus einfacher, da für die Stratifizierung nicht eine Vielzahl verschiedener Merkmale berücksichtigt werden müssen, sondern nur genau *eine* Zahl.

**Matchen** Mit Hilfe des PS teilt man die Teilnehmer in zwei Behandlungsgruppen auf, indem man jeweils einen Teilnehmer mit möglichst gleichem PS in die Gruppen A bzw. B aufnimmt. Man führt so eine „Quasi-Randomisierung“ durch. Dies ist mit dem PS weitaus einfacher durchzuführen als beim konventionellen Matchen, weil hier wiederum nur *eine* Zahl gematcht werden muss und nicht eine Vielzahl verschiedener Variablen.

**Gewichtung mit inverser Behandlungswahrscheinlichkeit** Wenn  $T$  die Zuweisung einer Behandlung ist ( $T = 1$  bedeutet Behandlung und  $T = 0$  bedeutet keine Behandlung) und  $Z$  der Propensity-Score, so ist die inverse Wahrscheinlichkeit der Behandlung  $\frac{T}{Z} + \frac{1-T}{1-Z}$ . Werden die Subjekte mit dieser inversen Wahrscheinlichkeit gewichtet, so erhält man eine künstliche Population, bei der die berücksichtigten Kovariablen unabhängig vom Behandlungsstatus sind. Das Confounding wird so für bekannte Confounder minimiert.

In einer Untersuchung über die Eignung der einzelnen Verfahren, für bekannte Kovariablen zu balancieren, wurde festgestellt, dass das Matchen und die Gewichtung mit der inversen Behandlungswahrscheinlichkeit nahezu alle Unregelmäßigkeiten zwischen den Behandlungsgruppen korrigieren konnten, während Stratifizierung und Regressions-Analyse mit dem PS als Kovariable ein wenig schlechter abschnitten [17].

Der Propensity-Score bietet also eine Reihe von Vorzügen: Er reduziert deutlich den Bias in nicht-randomisierten Studien. Einmal ausgerechnet kann er für eine Vielzahl unterschiedlicher Zielgrößen verwendet werden. Auch wenn die Zielgröße selten ist, hat er deutliche Vorteile gegenüber der Regressions-Analyse, da es dort zu viele Merkmale zu berücksichtigen gibt [8]. Nachteilig ist jedoch, dass beim Matchen mit dem Propensity-Score für unbeachtete Merkmale nicht balanciert wird (ebenso wenig wie bei der Regressions-Analyse). Außerdem lässt sich der PS schlecht anwenden, wenn nur wenige Teilnehmer zum Matchen zu Verfügung stehen, da es nun erheblich schwieriger wird, genau passende Matches zu finden. Vernachlässigt man dabei nicht matchbare Teilnehmer, so sinkt gleichzeitig die Power und die Anzahl der Studienteilnehmer [10]. Man kann weiterhin für die Berechnung des PS keine Kovariablen verwenden, die evtl. auch die Exposition, also die Therapie, beeinflussen. Zuletzt wird dem PS ein grundsätzlicher Nachteil in der Datenerhebung nachgesagt, da die PS-Studien die Strenge und die Disziplin randomisierter klinischer Studien vermissen ließen [18].

## 2.2 Auswahl der Daten

### 2.2.1 Auswahl der RCT-Studien

Als randomisiert-kontrollierte Studien wurden zunächst alle schon im Review von Møller [19] aufgeführten Studien übernommen. Dieses Review erfasste RCTs für den Zeitraum bis einschließlich Juli 2007, die Zielgrößen von On- bzw. Offpump-Bypass-Operationen untersuchten. Für den Zeitraum von Juli 2007 bis Januar 2011 wurde die Suche nach dem zugehörigen Studienprotokoll [20] wiederholt. Es wurde auf PubMed gesucht mit den Begriffen *Coronary Artery*

*Bypass, Off Pump[MeSh] OR off-pump[All Text] OR opcap[All Text] OR „beating heart”[All Text] OR op-cap[All Text] OR „off-pump”[All Text] OR octopus [All Text]*. Als Einschränkungen wurden gewählt: *Humans, Randomized Controlled Trial, Publication Date from 2007/05*. Es wurden nur solche Studien eingeschlossen, die On- und Off-Pump-Bypass-Operationen miteinander verglichen, in denen für diese Therapien randomisiert wurde und die von den gesuchten Merkmalen (siehe 2.2.3) mindestens eines aufführten. Doppelpublikationen wurden ausgeschlossen.

## **2.2.2 Auswahl der PS-Studien**

Für die Auswahl der PS-Studien wurde im Januar 2011 die Suche nach Propensity-Score gematchten, On- und Off-Pump-Technik bei Bypass-Operationen vergleichenden Studien wiederholt, die bereits für die Erstellung der MAPS-I-Studie durchgeführt worden war (siehe Studie von Kuß et al. [21] für Informationen zum Verfahren). Es wurden nur solche Studien eingeschlossen, die eine gematchte PS-Analyse beinhalteten und mindestens eines der unten genannten Merkmale enthielten. Doppelpublikationen wurden ausgeschlossen.

## **2.2.3 Extrahierte Daten**

Bei den RCT-Studien wurden sowohl für On- als auch für Off-Pump-Gruppen Daten über Alter, linksventrikuläre Ejektionsfraktion EF, Geschlecht, Diabetes, Hypertonie und Myokardinfarktgeschichte (MI) ausgelesen. Bei den PS-Studien wurden die gleichen Merkmale untersucht, jedoch, soweit gegeben, nicht nur die Daten der gematchten Population der PS-Studien (PSM) aufgenommen, sondern auch die der Grundgesamtheiten der PS-Studien (General). Hierbei muss in aller Deutlichkeit hervorgehoben werden, dass sich der Begriff „Grundgesamtheit” im Folgenden stets auf die in den verwendeten PS-Studien beschriebenen, also den jeweiligen PS-Studien zu Grunde liegenden *Klinikpopulationen* bezieht, aus der die Probanden für die gematchte PS-Studienpopulation rekrutiert wurden. Es handelt sich bei der „Grundgesamtheit” hier also nicht um den Durchschnitt aller existierenden Patienten, die mit On- oder Off-Pump-Bypass-Operation therapiert wurden.

Die Angaben zum Zeitraum der aufgetretenen Myokardinfarkte schwankten teilweise erheblich (zwischen ‘history of MI’ und einem Ereignis in den letzten 24 Stunden) und die gewonnenen Daten hätten keine Schlussfolgerungen zugelassen. Es wurden deshalb nur solche Angaben zur Myokardinfarktgeschichte übernommen, bei denen keinerlei Einschränkungen zum Zeitraum des Ereignisses angegeben wurde.

## **2.3 Hauptanalyse**

Ziel der Analysen war es, für jede Studienpopulation - RCT-Studienpopulation, gematchte PS-Studienpopulation und Grundgesamtheit - und für jedes der Merkmale einen Mittelwert  $\bar{x}$  bzw.

eine Häufigkeit  $\hat{p}$  mit Konfidenzintervall zu ermitteln, damit diese Werte hinsichtlich der externen Validität miteinander verglichen werden konnten. Dabei wurden in der Hauptanalyse zwei Schwerpunkte gelegt: ein Vergleich der Durchschnittswerte zwischen allen drei Studienpopulationen (im Weiteren als Between-Vergleich bezeichnet) und der Vergleich der Durchschnittswerte von Grundgesamtheit und gematchter PS-Studienpopulation derjenigen PS-Studien, die sowohl Daten für Grundgesamtheit und die gematchte Propensity-Score-Population angaben (Within-Vergleich). Es soll in Kapitel 2.3 zunächst die Bearbeitung der erhobenen Roh-Daten und anschließend die zu Grunde gelegten statistischen Modellierungen und Prozeduren der Hauptanalyse beschrieben werden.

Zur Überprüfung des Einflusses der zu Grunde gelegten statistischen Modellierung auf die Ergebnisse der Analyse wurde zusätzlich eine Sensitivitätsanalyse angefertigt, welche in Kapitel 2.4 beschrieben wird. Eine Übersicht über Unterschiede zwischen Haupt- und Sensitivitätsanalyse bietet Tabelle 1 auf Seite 11.

### 2.3.1 Einheitliche Lagemaße

Das Ziel der Datenerhebung und -bearbeitung war es, für jede Studie (und in dieser für jede der On- und Off-Pump-Gruppen) für die metrischen Merkmale Alter und Linksventrikuläre Ejektionsfraktion (EF) einen Mittelwert  $\bar{x}$  und für die binären Merkmale Geschlecht, Diabetes, Hypertonie und Myokardinfarktgeschichte die Häufigkeiten ihres Auftretens  $\hat{p}$  zu notieren. Da oft nicht die gesuchten, sondern andere Lagemaße angegeben waren, wurden sie mit Hilfe der in diesem Kapitel erläuterten Annahmen und Umrechnungen ermittelt bzw. imputiert.

**Zusammenführung von Subgruppen** In einigen Studien wurde die On- und/oder Off-Pump-Gruppe in weitere Subgruppen unterteilt. Zur Vereinheitlichung der gegebenen Daten und für die weitere Analyse wurden diese Subgruppen wieder zu einer On- bzw. einer Off-Pump zusammengefasst, indem Durchschnitte gebildet wurden.

**Imputation nach Hozo et al. [22]** War kein Mittelwert  $\bar{x}$ , jedoch Median  $\tilde{x}$  oder Range gegeben, so wurden nach Hozo et al. berechnet:

1. bei  $n > 25$

$$\bar{x} = \tilde{x} \tag{1}$$

2. bei  $n \leq 25$

$$\bar{x} = \frac{\min + 2 * \tilde{x} + \max}{4} \tag{2}$$

**Mittelwert und Median** Waren Angaben zu metrischen Merkmalen in Kategorien angegeben, wurde eine gleichmäßige Verteilung der Daten angenommen und der Median  $\tilde{x}$  errechnet. Da man von einer Gleichverteilung der Daten ausgeht, entspricht der Median dem Mittelwert. Die Formeln (3) und (4) vereinfachen die Rechnung.

**Tabelle 1:** Übersicht über statistische Modellierung, SAS-Prozeduren und Besonderheiten in der Berechnung der einzelnen Analysen.

	Hauptanalyse	Sensitivitätsanalyse
Between	Statistische Modellierung	Random-Effects-Metaanalyse
	SAS-Prozedur	PROC MIXED
	Besonderheiten	Gewichtung mit Fallzahlen Gewichtung mit inverser Varianz
Between	Statistische Modellierung	GLMM* mit Identitäts-Linkfunktion
	SAS-Prozedur	PROC GLIMMIX
	Besonderheiten	implizite Gewichtung im Modell Extremprävalenzen unproblematisch
Within	Statistische Modellierung	Random-Effects-Metaanalyse
	SAS-Prozedur	PROC MIXED
	Besonderheiten	Gewichtung mit Fallzahlen
Within	Statistische Modellierung	GLMM* mit Identitäts-Linkfunktion
	SAS-Prozedur	PROC GLIMMIX
	Besonderheiten	implizite Gewichtung im Modell Extremprävalenzen unproblematisch

\*Generalized linear mixed model (GLMM)

Beispiel: Aus der Angabe „60% der Patienten haben eine EF von  $> 50\%$ “ folgt: die untere Grenze des Intervalls (I)  $ug_I = 50$ , die obere Grenze des Intervalls  $og_I = 100$ , die Länge des Intervalls  $L_I = 50$  und die Besetzung des Intervalls  $B_I = 60$ .

Bei  $ug \neq 0$

$$\bar{x} = \tilde{x} = \frac{B_I - 50}{B_I} * L_I + ug_I \quad (3)$$

Bei  $ug = 0$

$$\bar{x} = \tilde{x} = \frac{og_I * 0,5}{B_I} \quad (4)$$

### 2.3.2 Fehlende, imputierte und vollständige Daten

Mit Hilfe der *PROC MEANS-Funktion* wurde in SAS für jede Studientyp und jedes Merkmal erhoben, wie viele Daten imputiert werden mussten, wie viele vollständig übernommen werden konnten und wie viele Daten fehlten.

### 2.3.3 Externe Validität - Between-Vergleich

**Annahmen** Ziel des systematischen Reviews war es, die externe Validität von gematchten PS-Studien und von RCTs zu beurteilen. Hierzu benötigt man die Merkmale des durchschnittlichen Bypass-Patienten zum Vergleich. Wie in 2.2.3 beschrieben, wird in dieser Arbeit mit dem Begriff „Grundgesamtheit“ diejenige Population bezeichnet, die den hier verwendeten PS-Studien zu Grunde liegt. Es wird jedoch angenommen, dass die durchschnittlichen Merkmale dieser ausgewählten Population in etwa dem wahren Durchschnitt aller Bypass-Patienten entsprechen und so als Vergleichspunkte für die externe Validität verwendet werden können.

Wie in Kapitel 2.3.1 beschrieben, mussten einige Mittelwerte der metrischen Daten imputiert werden. Die Imputation erfolgte gemäß Annahmen und Formeln nach Hozo et al. [22]. Hierbei wird angenommen, dass die zu Grunde liegenden Annahmen zur Imputation dieser Mittelwerte (die über den entsprechenden Median und den Range erfolgte) nach Hozo et al. so zutreffend sind, dass an dieser Stelle auf eine entsprechende Differenzierung und eine getrennte Analyse (mit bzw. ohne imputierte Daten) verzichtet werden kann.

**Statistische Modellierung** Bei der Analyse der Mittelwerte und Häufigkeiten kann man verschiedene statistische Modellierungen anwenden, die jeweils eigenen Grundannahmen über die Beschaffenheit der Daten folgen. Bei der Analyse der metrischen Merkmalen standen folgende Beobachtungsdaten zur Verfügung: der Mittelwert  $\bar{x}$ , die Fallzahl  $n$ , die Standardabweichung  $s$  und die Zugehörigkeit zu einer Studienpopulation (Grundgesamtheit, RCT oder gematchte PS-Studie). Gesucht wurde der Meta-Mittelwert in jeder Studienpopulation (also z.B. das durchschnittliche Alter aller RCT-Studienteilnehmer). Der Meta-Mittelwert sollte für die Präzision der Schätzung für jede einzelne Beobachtung gewichtet werden. Dabei mussten

zwei Probleme in der Analyse gelöst werden. Erstens lieferten manche PS-Studien zwei Beobachtungen: eine für die Grundgesamtheit und eine für die gematchte Studienpopulation. Es musste sichergestellt werden, dass diese Populationen, bei der die gematchten Probanden aus der Grundgesamtheit rekrutiert werden, in der Analyse statistisch unabhängig waren. Zweitens ergaben sich Schwierigkeiten bei der Gewichtung (Näheres im folgenden Abschnitt *Gewichtung*). Wie sich aus Tabelle 1 auf Seite 11 entnehmen lässt, wurde zur Lösung dieser Probleme in der Hauptanalyse der metrischen Daten das sog. Random-Effects-Modell (REM) zu Grunde gelegt. Dies ist ein Meta-Regressions-Modell mit zufälligem Effekt, durch das die korrekte Modellierung zweier korrelierter Populationen sichergestellt wird.

Bei der Analyse der binären Daten standen als Beobachtungen die Prävalenz des jeweiligen Merkmals, sowie die Zugehörigkeit zu einer Studienpopulation (Grundgesamtheit, RCT oder gematchte Ps-Studie) zur Verfügung. Gesucht wurde die Meta-Prävalenz des Merkmals in jeder Studienpopulation, wiederum - wie bei den metrischen Merkmalen - adjustiert für die Präzision der Schätzung für die einzelne Beobachtung. Hier musste ebenfalls der Tatsache Rechnung getragen werden, dass manche PS-Studien zwei Populationen beschrieben, wobei die gematchte Population aus der Grundgesamtheit hervorging. Es wurde ein Generalized linear mixed model (GLMM) mit Identitäts-Linkfunktion verwandt. Das GLMM ist ein statistisches Modell, das die Klasse der Generalized linear models (GLM) erweitert, indem es einen normalverteilten zufälligen Effekt zusätzlich einschließt. Während das GLM davon ausgeht, dass die zu analysierenden Daten unkorreliert sind, wird beim GLMM eine Korrelation berücksichtigt. Zudem werden Grundgesamtheit und die aus ihr rekrutierte gematchte Population als statistisch unabhängig voneinander bewertet. Die Link-Funktion stellt einen Zusammenhang zwischen linearem Prädiktor und Mittelwert der angenommenen Verteilung dar [23–25].

**Gewichtung ( $w_i$ )** Es wurde für metrische Merkmale mit der Fallzahl gewichtet, da bei Gewichtung mit der inversen Varianz die Ergebnisse durch die teilweise fehlerhaft angegebenen Ursprungsdaten aus den Studien (siehe 3.2.1 auf Seite 20) bzw. durch die große Zahl der durch Imputation ungenauen Standardabweichungen verfälscht worden wären (zu Imputation der Standardabweichungen, siehe Kapitel 2.4.1 auf Seite 15). Zudem hätten kleine Studien mit einer kleinen Standardabweichung ein unverhältnismäßig größeres Gewicht gehabt, als große Studien mit großen Standardabweichung. Es wurde die Fallzahl jeder Studie durch die Gesamtfallzahl der individuellen Beobachtungen geteilt (Alter: 428930, EF: 326185), und anschließend multipliziert mit der Anzahl der Studien, die in die jeweilige Analyse eingehen.

$$w_{iOn,Off} = (n_{On} + n_{Off}) * \frac{n_{AnzahlStudien}}{Gesamtfallzahl} \quad (5)$$

Beispiel: Es wurden bei den RCT-Studien 140 Studien für die Analyse des Alters und 106 für die Analyse des EF eingeschlossen. Die Berechnung der Gewichts  $w_{iOn,Off}$  für das mittlere Alter der RCT-Studienpopulation erfolgt in diesem Falle wie in Formel (6):

$$w_{iOn,Off} = (n_{On} + n_{Off}) * \frac{140}{428930} \quad (6)$$

Für binäre Merkmale wird im General Linear Mixed Model implizit gewichtet (umgesetzt von der SAS-Prozedur PROC GLIMMIX). Es war nicht nötig, manuell ein Gewicht zu errechnen.

Es wurden neben den  $\bar{x}_{MA}$  bzw.  $\hat{p}_{MA}$  für jede einzelne Studienpopulation auch die Differenzen dieser Lagemaße zwischen den Studienpopulationen errechnet. Die Differenzen wurden in den SAS-Prozeduren PROC MIXED bzw. PROC GLIMMIX implizit berechnet. Sie erlauben einen direkten Vergleich der Merkmalsausprägungen in den Populationen und eine Aussage über die Signifikanz der gefundenen Unterschiede. Weitere Informationen zu den mathematischen Hintergründen der genannten SAS-Prozeduren finden sich in den Arbeiten von Hedeker [25] bzw. SAS Institute Inc. [26].

### 2.3.4 Externe Validität - Within-Vergleich

Um festzustellen, ob sich die durchschnittlichen Merkmale der Grundgesamtheit durch Matchen grundsätzlich ändern, wurde der Within-Vergleich angestellt. Dafür wurden nur diejenigen PS-Studien einbezogen, die Angaben sowohl zur Grundgesamtheit als auch zu gematchten Probanden machten. Es wurde für jede Studie jeweils  $\bar{x}_{On,Off}$  bzw.  $\hat{p}_{On,Off}$  von Grundgesamtheit und gematchter Studienpopulation ermittelt und die Differenzen der Mittelwerte gebildet. Es erfolgte die Analyse mit Modellen und SAS-Prozeduren analog zum Between-Vergleich, eine Übersicht bietet wiederum Tabelle 1 auf Seite 11. Wie auch beim Between-Vergleich wurden neben den  $\bar{x}_{MA}$  und  $\hat{p}_{MA}$  für Grundgesamtheit und gematchte PS-Studienpopulation die Differenz der beiden Erwartungswerte  $\Delta_{General-PSM}$  gebildet, sodass eine Aussage zu Signifikanz und Ausprägungen der Merkmalsunterschiede getroffen werden kann. Die Berechnungen von  $\bar{x}_{MA}$ ,  $\hat{p}_{MA}$  und  $\Delta_{General-PSM}$  erfolgten mit der SAS-Prozedur PROC MIXED für metrische Daten und PROC GLIMMIX für binäre Daten.

## 2.4 Sensitivitätsanalyse

Mit Hilfe der Sensitivitätsanalyse sollte eruiert werden, ob sich die Ergebnisse der Hauptanalyse auf unterschiedliche Charakteristika der Studienpopulationen zurückführen lassen, oder ob die Ergebnisse von der zu Grunde gelegten statistischen Modellierung abhängig sind. Dazu wurde die Between-Analyse der externen Validität unter anderen Grundannahmen und Vereinfachungen wiederholt. Analog zum Kapitel 2.3 sollen zunächst die nötigen Modifikationen der erhobenen Roh-Daten und anschließend die zu Grunde gelegten Annahmen, statistischen Modellierungen und Prozeduren beschrieben werden.

### 2.4.1 Einheitliche Lage- und Streuungsmaße

Da in der Sensitivitätsanalyse mit der inversen Varianz gewichtet wurde, musste - *zusätzlich* zu einheitlichen Lagemaßen (siehe Kapitel 2.3.1) - für die metrischen Merkmale Alter und EF ein einheitliches Streuungsmaß (Standardabweichung  $s$ ) berechnet werden. Dabei wurden die in diesem Kapitel beschriebenen Vereinfachungen und Annahmen verwandt.

**Zusammenführung von Subgruppen** Der Durchschnitt der Standardabweichungen  $\bar{s}$  wurde berechnet nach Formel (7):

$$\bar{s} = \sqrt{\frac{n_1 * s_1^2 + n_2 * s_2^2}{n_1 + n_2}} \quad (7)$$

**Imputation nach Börgermann et al. [27] und Hozo et al. [22]** Fehlte bei metrischen Merkmalen die Standardabweichung  $s$  bzw. wurde nur ein Minimum oder ein Maximum des Ranges angegeben, wurden ausgehend von der Hallenser 3CAB-Studie [27] die unten aufgeführten Minima bzw. Maxima für den Range (R) der metrischen Merkmale Alter und EF imputiert und mit den Formeln (8) - (10) nach Hozo u. a. [22] die Standardabweichung berechnet.

Alter [Jahre]:  $min = 40$   $max = 85$   $R = max - min = 45$

EF [%]:  $min = 10$   $max = 90$   $R = max - min = 80$

1. bei  $n \leq 25$ :

$$s = \sqrt{\frac{\frac{(min - 2*\bar{x} + max)^2}{4} + (max - min)^2}{12}} * 100 \quad (8)$$

2. bei  $25 < n \neq 70$

$$s = \frac{R}{4} \quad (9)$$

3. bei  $70 < n$

$$s = \frac{R}{6} \quad (10)$$

War anstelle der Standardabweichung  $s$  der Standardfehler  $\sigma_{\bar{x}}$  gegeben, wurde die Standardabweichung berechnet:

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} \quad (11)$$

War anstelle der Standardabweichung  $s$  der Interquartilabstand  $I_{50}$  gegeben, wurde angenommen, dass dieser aus einer Normalverteilung stammt. In der Normalverteilung liegen 50% der Werte zwischen  $\mu - 0,675 * \sigma$  und  $\mu + 0,675 * \sigma$ . Es folgt daraus:

$$s = \frac{I_{50}}{2 * 0,675} \quad (12)$$

War das 95%-Konfidenzintervall (95%KI) und der Mittelwert  $\bar{x}$  gegeben, so wurde die Standardabweichung  $s$  nach Formel (13) berechnet:

$$s = \frac{\bar{x} - ug_{95\%KI}}{1,96} * \sqrt{n} \quad (13)$$

#### 2.4.2 Externe Validität - Between-Vergleich

**Imputation** In der Sensitivitätsanalyse wurden nicht nur Mittelwerte, sondern auch Standardabweichungen nach Hozo et al. [22] bzw. auf Datengrundlage der 3CAB-Studie [27] imputiert. Um festzustellen, welchen Einfluss die Imputation von Mittelwert und Standardabweichung auf die Ergebnisse hat, wurden für die metrischen Merkmale Alter und EF deshalb zwei Berechnungsgrundlagen gewählt: zum einen der Datensatz *mit* imputierten Daten und zum anderen derjenige mit *ausschließlich* vollständigen, also nicht imputierten Daten.

**Statistische Modellierung** Sowohl die Analyse der metrischen, als auch der binären Daten wurden unter Annahme des Random-Effects-Modells durchgeführt (Erläuterungen zu Modellen siehe Kapitel 2.3.3). Im REM geht man von einer Normalverteilung der metrischen Daten aus und von einer Beta-Binomial-Verteilung der binären Daten.

**Gewichtung** Um den Effekt unterschiedlicher Gewichtungsmöglichkeiten auf die Ergebnisse zu untersuchen, wurde in der Sensitivitätsanalyse mit der inversen Varianz gewichtet (Ermittlung des Gewichtes über inverse Varianz siehe Formeln (14) bis (19)).

$$se_{\bar{x}_{On,Off}} = \frac{s_{On,Off}}{\sqrt{n_{On} + n_{Off}}} \quad (14) \quad se_{\hat{p}_{On,Off}} = \sqrt{\frac{(\hat{p}_{On,Off}) * (1 - \hat{p}_{On,Off})}{n_{On} + n_{Off}}} \quad (15)$$

$$Var_{\bar{x}_{On,Off}} = se_{\bar{x}_{On,Off}}^2 \quad (16) \quad Var_{\hat{p}_{On,Off}} = se_{\hat{p}_{On,Off}}^2 \quad (17)$$

$$w_{iOn,Off} = \frac{1}{Var_{\bar{x}_{On,Off}}} \quad (18) \quad w_{iOn,Off} = \frac{1}{Var_{\hat{p}_{On,Off}}} \quad (19)$$

**Extremprävalenzen** Da in der Sensitivitätsanalyse der binären Daten mit der inversen Varianz gewichtet wurde und bei Prävalenzen  $\hat{p}_{On,Off}$  von 0% bzw. 100% durch 0 geteilt worden wäre, mussten Extremprävalenzen korrigiert werden. Wenn  $\hat{p}_{On,Off} = 0$  oder  $\hat{p}_{On,Off} = 1$  gilt:

$$\hat{p}_{On,Off} = \frac{(x_{On} + x_{Off}) + 0,01}{(n_{On} + n_{Off}) + 0,02} \quad (20)$$

Dabei stellt  $x$  die Anzahl der Merkmalsträger und  $n$  die Gesamtzahl der Patienten einer Gruppe dar. In der Hauptanalyse der binären Daten wurden Extremprävalenzen in der GLIMMIX-Prozedur adäquat berücksichtigt.

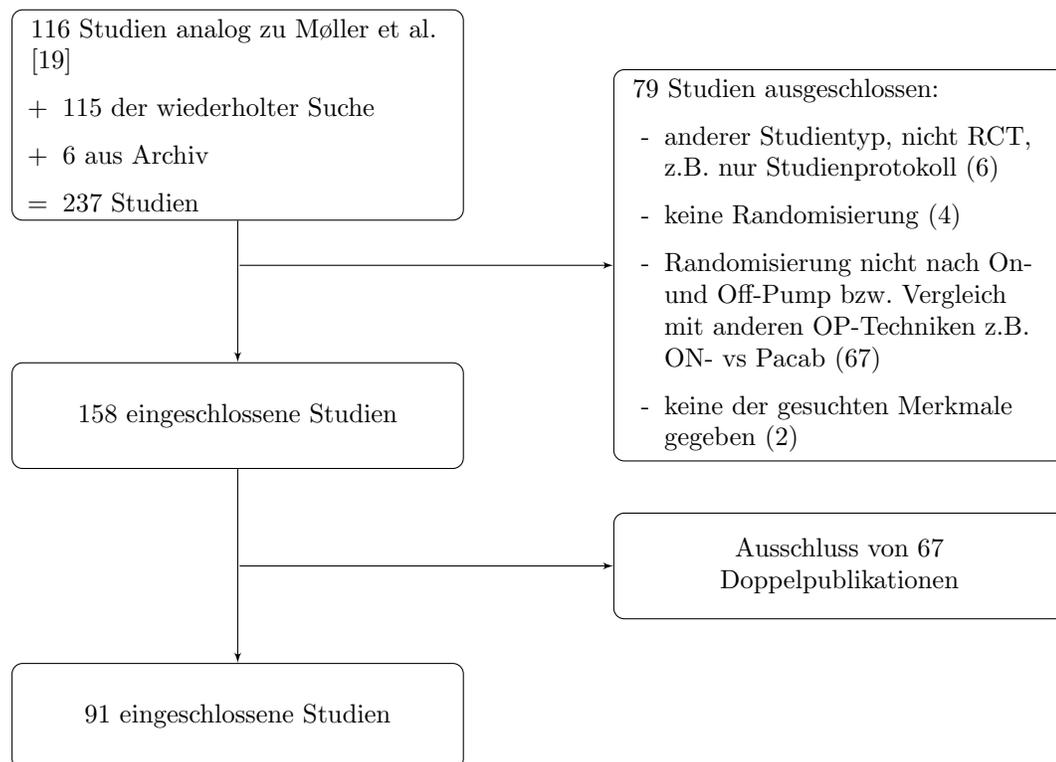
Ebenso wie in der Hauptanalyse wurden auch hier die Differenzen der Lagemaße zwischen den Studienpopulationen errechnet. Diese wurden sowohl für metrische als auch für binäre Merkmale in der SAS-Prozedur PROC MIXED implizit berechnet.

### 3 Ergebnisse

#### 3.1 Ausgewählte Studien

##### 3.1.1 RCT-Studien

Als randomisiert-kontrollierte Studien wurden 116 Studien aus Møller et al. [19] übernommen und nochmals 115 aus der wiederholten Suche nach entsprechendem Protokoll [20] gefunden. Des Weiteren fanden sich im Archiv 6 Studien, die die Suchkriterien erfüllten; insgesamt ergab die Suche 237 Studien. Nach Ausschluss von Studien aufgrund der in Kapitel 2.2.1 genannten Kriterien verblieben 158 Studien. Nach Ausschluss von 67 Doppelpublikationen verblieben 91 zu analysierende Studien. In Abbildung 2 ist die Studiensuche schematisch dargestellt. Eine alphabetische Auflistung aller eingeschlossenen RCT-Studien findet sich ab Seite 48. Im folgenden Text werden diese RCT-Studien mit [RCTxx] referenziert.

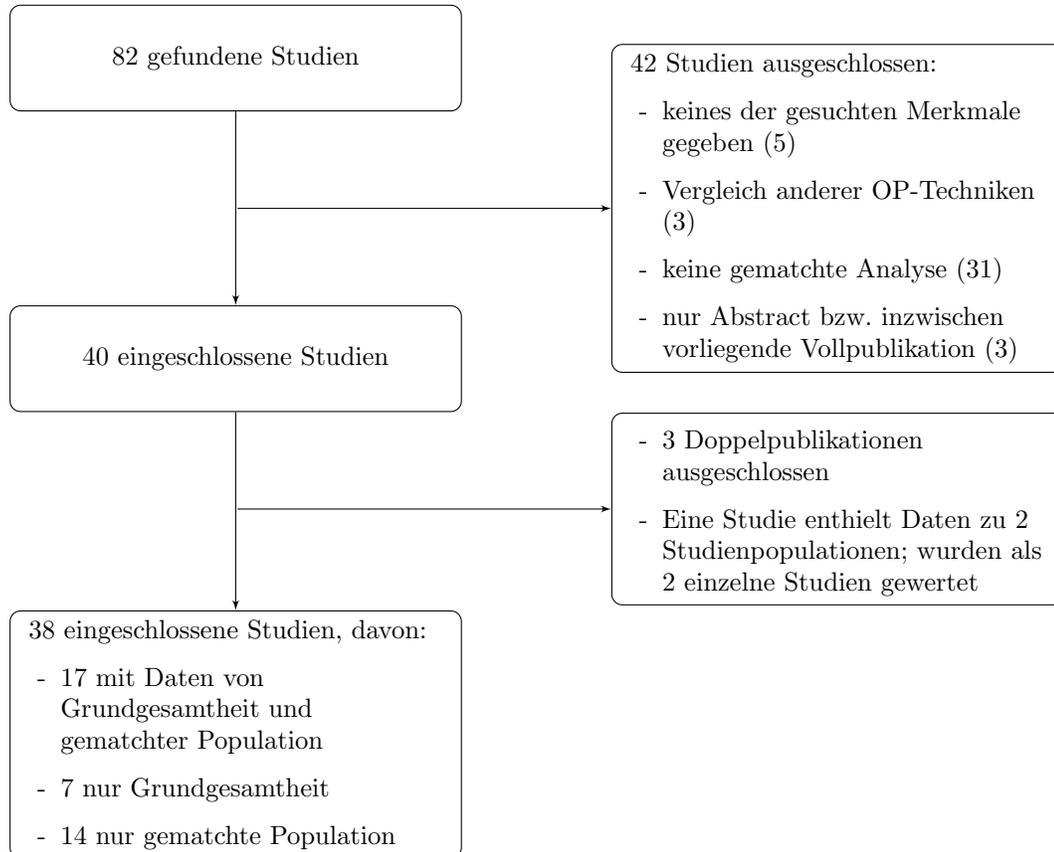


**Abbildung 2:** Fließdiagramm der Studiensuche zu randomisiert-kontrollierten Studien.

##### 3.1.2 PS-Studien

Als gematchte Propensity-Score-Studien wurden nach Suche in Pubmed nach dem in der Arbeit von Kuß et al. [21] beschriebenen Protokoll 82 Studien gefunden. Nach Ausschluss von 42 Studien aufgrund der in Kapitel 2.2.2 genannten Kriterien verblieben 40 Studien. Da drei

Doppelpublikation ausgeschlossen wurden und eine Studie Angaben zu 2 Studienpopulationen enthielt, ergaben sich als Suchergebnis 38 eingeschlossene Studien. Dabei konnten bei 17 Studien Daten sowohl von der Grundgesamtheit als auch von gematchten Gruppen, bei 7 Studien nur von der Grundgesamtheit und bei weiteren 14 Studien nur von gematchten Gruppen gewonnen werden. In Abbildung 3 ist die Studiensuche schematisch dargestellt. Eine alphabetische Auflistung aller eingeschlossenen Propensity-Score-Studien findet sich ab Seite 57. Im folgenden Text werden diese Propensity-Score-Studien mit [PSxx] referenziert.

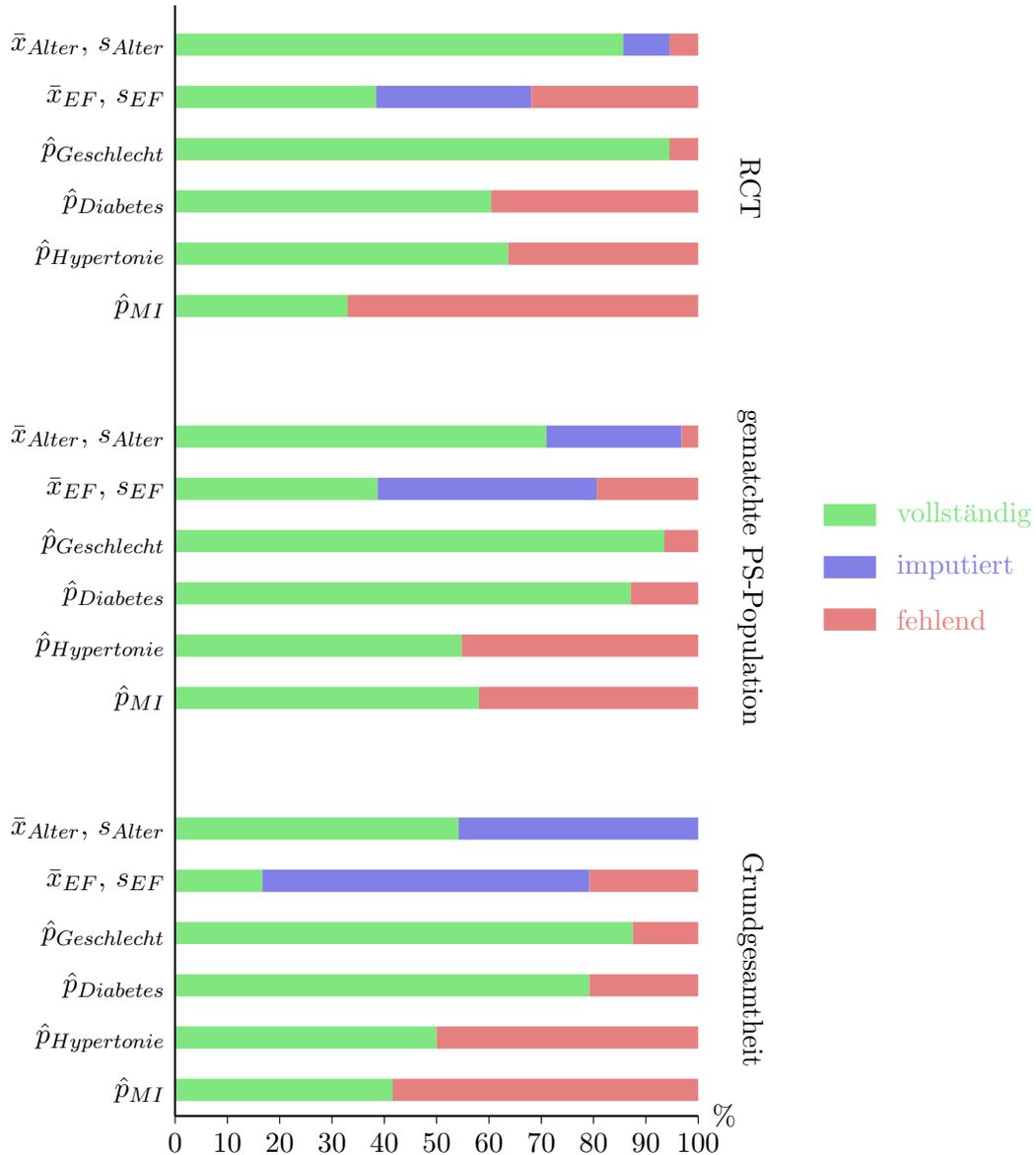


**Abbildung 3:** Fließdiagramm der Studiensuche zu gematchten Propensity-Score-Studien.

### 3.2 Extrahierte Daten

Um einen Überblick über die verfügbaren Daten zu erlangen, wurde zunächst erhoben, wie viele der eingeschlossenen Studien überhaupt Angaben zu den gesuchten Merkmalen enthielten, wie viele der Daten übernommen werden konnten, und in welchem Ausmaß imputiert werden musste, um für metrische Merkmale einen Mittelwert  $\bar{x}$  und eine Standardabweichung  $s$  zu erhalten. Eine ausführliche Tabellen-Übersicht der Merkmalsausprägungen findet sich auf Seite 61.

Die Abbildung 4 auf Seite 19 verdeutlicht, dass Angaben zum Geschlecht der Patienten in den meisten Studien vorhanden waren, während Angaben zu den Merkmalen Diabetes, Hypertonie und Myokardinfarktgeschichte besonders in den RCT-Studien oft fehlten. Besonders klein war der Anteil der übernommenen Daten bei der Myokardinfarktgeschichte, da hier deutliche Einschränkungen in der Auswahl vorgenommen wurden (siehe Kapitel 2.2.3).



**Abbildung 4:** Dargestellt sind die Anteile der vollständig übernommenen, der imputierten und der fehlenden Daten der Probanden der Grundgesamtheit sowie der RCT- bzw. der PS-Studien in den untersuchten Studien.

### 3.2.1 Umgang mit unplausiblen und problematischen Daten

Bei der Betrachtung der extrahierten Daten fielen bei einigen Studien Unstimmigkeiten auf. Oft war die Standardabweichung der metrischen Merkmale Alter und EF mit unrealistisch kleinen Zahlen angegeben. Die Autoren der jeweiligen Studien wurden gebeten zur Klärung der Unstimmigkeiten beizutragen.

**Motallebzadeh et al. [RCT53]** Die für diese Arbeit verwendete Standardabweichung  $s$  für das Alter weicht vom Paper ab, weil dort die Standardabweichung  $s$  mit dem Standardfehler  $\sigma_{\bar{x}}$  vertauscht wurde. Der Autor selbst korrigierte diese Angaben.

**Rainio et al. [RCT70]** Die un plausible Standardabweichung  $s$  des Alters wurde auf Nachfrage vom Autor korrigiert.

**Dorman et al. [RCT21], Meco et al. [PS26]** Die Merkmale Alter und EF in Dorman et al. bzw. nur das Alter in Meco et al. wurden für die Auswertung der Daten entfernt, da Unstimmigkeiten bezüglich der Standardabweichung  $s$  nicht befriedigend von den Autoren geklärt werden konnten und keine weiteren Angaben wie z.B. der Range gegeben waren, aus denen  $s$  hätte berechnet werden können.

**Malik et al. [RCT43]** Die Standardabweichung  $s$  der linksventrikulären Ejektionsfraktion ist unplausibel. Da der Autor nicht zur Klärung des Problems beitragen konnte, wurde die Standardabweichung aus den Angaben zum Range berechnet.

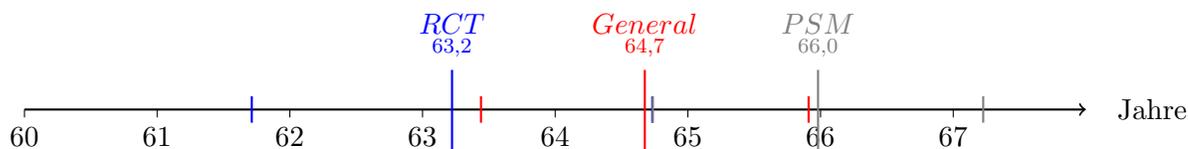
## 3.3 Hauptanalyse

### 3.3.1 Externe Validität - Between-Vergleich

Die statistische Auswertung der gewichteten Meta-Mittelwerte  $\bar{x}_{MA}$  und Meta-Häufigkeiten  $\hat{p}_{MA}$  (Metaanalyse (MA)) ergab die in den Tabellen 2 und 3 auf Seite 23 und 24 aufgeführten Werte. Dort sind ebenfalls die errechneten Differenzen der Mittelwerte bzw. der mittleren Häufigkeiten der Studienpopulationen gelistet. Ab Seite 62 findet sich eine ausführliche Übersicht der Ausprägungen der Merkmale in den einzelnen Studien.

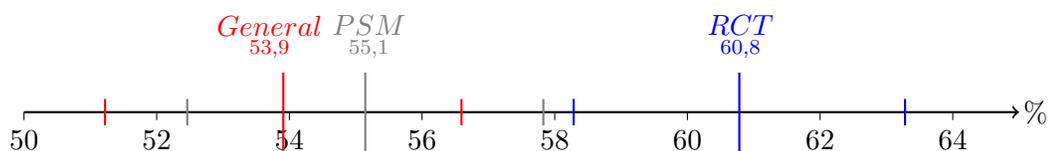
Im Folgenden werden zunächst die errechneten  $\bar{x}_{MA}$  und  $\hat{p}_{MA}$  jeder Studienpopulation mit Konfidenzintervall für jedes Merkmal dargestellt. Im Abschnitt 3.4 folgen die Ergebnisse der Sensitivitätsanalyse (Unterschiede der statistischen Modellierung, siehe Kapitel 2.3.3).

**Alter** Im Schnitt sind die Probanden der gematchten Propensity-Score-Studien signifikant älter als die der Grundgesamtheit ( $\Delta_{General-PSM} = -1,3$ ; [95%KI :  $-1,8$ ;  $-0,8$ ]) und die der RCTs ( $\Delta_{PSM-RCT} = 2,8$ ; [95%KI :  $0,8 - 4,7$ ]). Der hier festgestellte Altersunterschied von 1,5 Jahren zwischen Grundgesamtheit und RCT ( $\Delta_{General-RCT} = 1,5$ ; [95%KI :  $-0,5$ ;  $4,7$ ]) kann auf Grund mangelnder Signifikanz zwar nicht generalisiert werden, zeigt jedoch einen gewissen Trend an.



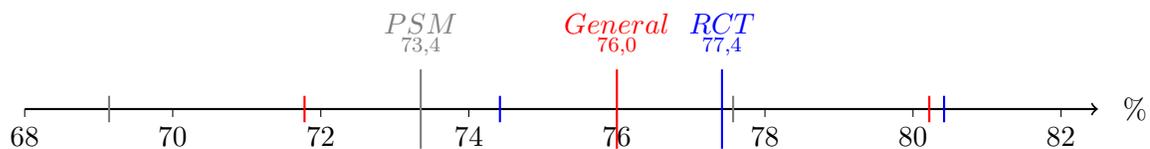
**Abbildung 5:** Meta-Mittelwerte  $\bar{x}_{MA}$  des Alters der drei Studienpopulationen mit ihrem 95%-Konfidenzintervall.

**EF** Bei der linksventrikulären Ejektionsfraktion lassen sich deutliche Unterschiede zwischen den Populationen feststellen. Sowohl die Teilnehmer der RCTs als auch die der gematchten PS-Studien haben signifikant höhere Ejektionsfraktionen als die Grundgesamtheit, sind also in diesem Aspekt gesünder. Dabei ist hervorzuheben, dass die Differenz zwischen Grundgesamtheit und gematchten PS-Studien deutlich kleiner ist ( $\Delta_{General-PSM} = -1,2$ ; [95%KI :  $-2,3$ ;  $-0,2$ ]) als die Differenz zwischen Grundgesamtheit und RCT-Studien ( $\Delta_{General-RCT} = -6,9$ ; [95%KI :  $-10,5$ ;  $3,2$ ]). Die Differenz von 6,9 Prozentpunkten ist als erheblich zu betrachten.



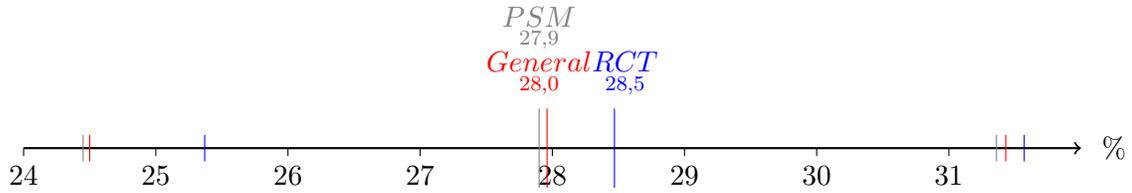
**Abbildung 6:** Meta-Mittelwerte  $\bar{x}_{MA}$  der linksventrikulären Ejektionsfraktion der drei Studienpopulationen mit ihrem 95%-Konfidenzintervall.

**Geschlecht** Der Anteil der männlichen Probanden ist in den RCTs geringfügig und nicht signifikant größer als in der Grundgesamtheit ( $\Delta_{General-RCT} = -1,4$ ; [95%KI :  $-6,6$ ;  $3,8$ ]), in den gematchten PS-Studien dagegen signifikant kleiner ( $\Delta_{General-PSM} = 2,6$ ; [95%KI :  $2,2$ ;  $3,1$ ]). Der große Unterschied zwischen Probanden in gematchten PS-Studien und RCTs ( $\Delta_{PSM-RCT} = -4,1$ ; [95%KI :  $-9,2$ ;  $1,1$ ]) lässt sich auf Grund des Konfidenzintervalls nicht generalisieren.



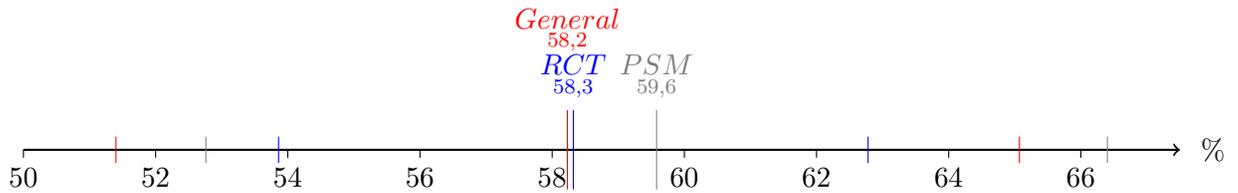
**Abbildung 7:** Meta-Häufigkeiten  $\hat{p}_{MA}$  des Anteils der männlichen Patienten in den drei Studienpopulationen mit ihrem 95%-Konfidenzintervall.

**Diabetes** Betrachtet man den Anteil der Diabetiker in den verschiedenen Studienpopulationen, so sind die festgestellten Unterschiede marginal und nicht signifikant. Wie sich aus Tabelle 3 ersehen lässt, sind die Differenzen mit höchstens 0,5 Prozentpunkten vernachlässigbar klein.



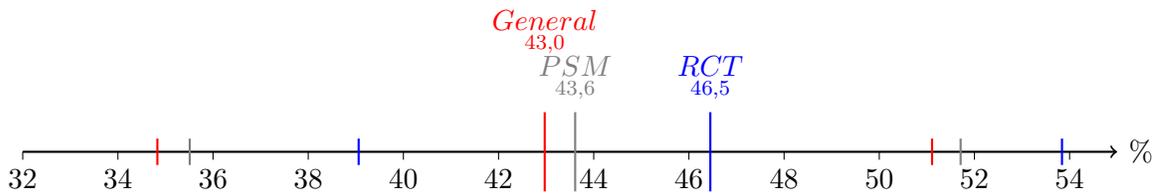
**Abbildung 8:** Meta-Häufigkeiten  $\hat{p}_{MA}$  des Anteils der Diabetiker in % in den drei Studienpopulationen mit ihrem 95%-Konfidenzintervall.

**Hypertonie** Es ließen sich ebenfalls nur leichte Unterschiede zwischen den Populationen feststellen. Die Teilnehmer der gematchten PS-Studien wiesen zwar im Schnitt eine höhere Wahrscheinlichkeit auf, unter Hypertonie zu leiden, als die der Grundgesamtheit ( $\Delta_{General-PSM} = -1,4$ ; [95%KI : -3,1; 0,4]) bzw. der RCTs ( $\Delta_{PSM-RCT} = 1,3$ ; [95%KI : -6,9; 9,4]), jedoch sind diese Unterschiede mit höchstens 1,4 Prozentpunkten erstens nicht groß und zweitens auf Grund ihres Konfidenzintervalls auch nicht signifikant. Der Unterschied zwischen Grundgesamtheit und den RCTs ist mit 0,1 Prozentpunkten vernachlässigbar klein ( $\Delta_{General-RCT} = -0,1$ ; [95%KI : -8,3; 8,1]).



**Abbildung 9:** Meta-Häufigkeiten  $\hat{p}_{MA}$  des Anteils der Hypertoniker in den drei Studienpopulationen mit ihrem 95%-Konfidenzintervall.

**Myokardinfarktgeschichte** Sowohl in gematchten PS-Studien als auch in RCTs haben die Probanden eine *höhere* Wahrscheinlichkeit, in der Vergangenheit einen Myokardinfarkt erlitten zu haben. Sie sind bezüglich dieses Merkmals also *weniger* gesund als die Grundgesamtheit. So liegt die Häufigkeit bei RCT-Probanden mit 46,5% um 3,5 Prozentpunkte höher als bei der Grundgesamtheit, jedoch ist keine der in errechneten Differenzen signifikant und lässt sich so nicht generalisieren (siehe Konfidenzintervalle in Tabelle 3).



**Abbildung 10:** Meta-Häufigkeiten  $\hat{p}_{MA}$  des Anteils der Patienten mit Myokardinfarktgeschichte in % in den drei Studienpopulationen mit ihrem 95%-Konfidenzintervall.

**Tabelle 2:** Between-Analyse der metrischen Merkmale. Aufgeführt sind die errechneten Meta-Mittelwerte  $\bar{x}_{MA}$  (oberer Abschnitt) bzw. die Differenzen  $\Delta$  der  $\bar{x}_{MA}$  zwischen den drei Studienpopulationen (unterer Abschnitt) in Haupt- und Sensitivitätsanalyse.

		Hauptanalyse				Sensitivitätsanalyse						
		I+V		V		I+V		V				
		$n$	$\bar{x}_{MA}$	95%KI	$n$	$\bar{x}_{MA}$	95%KI	$n$	$\bar{x}_{MA}$	95%KI		
Alter [Jahre]	General	24	64,7	63,4 ; 65,9	24	65,1	63,9 ; 66,4	13	65,5	64,1 ; 66,9		
	PSM	30	66,0	64,7 ; 67,2	30	66,1	64,9 ; 67,4	22	66,1	64,7 ; 67,5		
	RCT	86	63,2	61,7 ; 64,7	86	63,2	62,3 ; 64,0	78	63,2	62,4 ; 64,0		
EF [%]	General	19	53,9	51,2 ; 56,6	19	53,8	51,0 ; 56,6	4	50,7	45,8 ; 55,6		
	PSM	25	55,1	52,5 ; 57,8	25	54,9	52,1 ; 57,7	12	50,6	45,7 ; 55,5		
	RCT	62	60,8	58,3 ; 63,3	62	60,5	58,5 ; 62,6	35	58,6	55,7 ; 61,5		
Alter [Jahre]	General-PSM	$\Delta$	-1,3	-1,8 ; -0,8	$\Delta$	-1,0	-1,1 ; -0,9	$\Delta$	-0,6	-0,8 ; -0,5	p-Wert	<,0001
	General-RCT		1,5	-0,5 ; 3,4		2,0	0,5 ; 3,5		2,3	0,6 ; 3,9		0,0063
	PSM-RCT		2,8	0,8 ; 4,7		3,0	1,4 ; 4,5		2,9	1,3 ; 4,6		0,0005
EF [%]	General-PSM		-1,2	-2,3 ; -0,2		-1,2	-1,4 ; -0,9		0,1	-0,3 ; 0,5		0,6703
	General-RCT		-6,9	-10,5 ; -3,2		-6,8	-10,2 ; -3,3		-7,9	-13,7 ; -2,2		0,0070
	PSM-RCT		-5,6	-9,3 ; -2,0		-5,6	-9,1 ; -2,1		-8,0	-13,7 ; -2,3		0,0063

imputierte und vollständige Daten (I+V), ausschließlich vollständige Daten (V).

**Tabelle 3:** Between-Analyse der binären Merkmale. Aufgeführt sind die errechneten Meta-Häufigkeiten  $\hat{p}_{MA}$  (oberer Abschnitt) bzw. die Differenzen  $\Delta$  der  $\hat{p}_{MA}$  zwischen den drei Studienpopulationen (unterer Abschnitt) in Haupt- und Sensitivitätsanalyse.

		Hauptanalyse			Sensitivitätsanalyse		
		$n$	$\hat{p}_{MA}$	95%KI	$n$	$\hat{p}_{MA}$	95%KI
<i>Geschl.</i> [%]	General	21	76,0	71,8 ; 80,2	21	76,03	72,23 ; 79,82
	PSM	29	73,4	69,1 ; 77,6	29	73,45	69,66 ; 77,24
	RCT	86	77,4	74,4 ; 80,4	86	77,79	75,10 ; 80,47
<i>Diab.</i> [%]	General	19	28,0	24,5 ; 31,6	19	28,06	24,93 ; 31,19
	PSM	27	27,9	24,2 ; 31,3	27	27,71	24,59 ; 30,82
	RCT	55	28,5	24,8 ; 31,1	55	27,71	24,95 ; 30,47
<i>Hypert.</i> [%]	General	12	58,2	51,4 ; 65,1	12	58,20	52,37 ; 64,03
	PSM	17	59,6	52,8 ; 66,4	17	59,58	53,77 ; 65,40
	RCT	58	58,3	53,9 ; 62,8	58	58,51	54,75 ; 62,27
<i>MI</i> [%]	General	10	43,0	34,8 ; 51,1	10	43,02	36,15 ; 49,90
	PSM	18	43,6	35,5 ; 51,7	18	43,63	36,79 ; 50,48
	RCT	29	46,5	39,1 ; 53,8	29	46,64	40,48 ; 52,80

		$\Delta$	95%KI	p-Wert	$\Delta$	95%KI	p-Wert
<i>Geschl.</i> [%]	General-PSM	2,6	2,2 ; 3,1	<,0001	2,6	2,2 ; 3,0	<,0001
	General-RCT	-1,4	-6,6 ; 3,8	0,5627	-1,8	-6,4 ; 2,9	0,4577
	PSM-RCT	-4,1	-9,2 ; 1,1	0,1135	-4,3	-9,0 ; 0,3	0,0671
<i>Diab.</i> [%]	General-PSM	0,06	-1,0 ; 1,1	0,9016	0,4	-0,5 ; 1,2	0,4171
	General-RCT	0,5	-5,2 ; 4,1	0,8134	0,3	-3,8 ; 4,5	0,8710
	PSM-RCT	0,5	-5,2 ; 4,1	0,7919	-0,006	-4,2 ; 4,1	0,9979
<i>Hypert.</i> [%]	General-PSM	-1,4	-3,1 ; 0,4	0,1027	-1,4	-2,8 ; -0,006	0,0491
	General-RCT	-0,1	-8,3 ; 8,1	0,9801	-0,3	-7,2 ; 6,6	0,9301
	PSM-RCT	1,3	-6,9 ; 9,4	0,7253	1,1	-5,9 ; 8,0	0,7613
<i>MI</i> [%]	General-PSM	-0,6	-2,2 ; 0,9	0,3624	-0,6	-1,8 ; 0,6	0,3282
	General-RCT	-3,5	-14,5 ; 7,5	0,4687	-3,6	-12,8 ; 5,6	0,4427
	PSM-RCT	-2,8	-13,8 ; 8,1	0,5502	-3,0	-12,2 ; 6,2	0,5225

Geschlecht (Geschl.), Diabetes (Diab.), Hypertonie (Hypert.)

### 3.3.2 Externe Validität - Within-Vergleich

Um einschätzen zu können, ob und wie sich beim Matchen die Merkmalsausprägung einer Population in einer PS-Studie verändert, wurden  $\bar{x}_{MA}$  und  $\hat{p}_{MA}$  für Grundgesamtheit und gematchte PS-Studienpopulation derjenigen PS-Studien miteinander verglichen, die sowohl Angaben zur Grundgesamtheit, als auch zu den Probanden der Studie *nach* Matchen enthielten. In Tabelle 4 auf Seite 26 werden sowohl  $\bar{x}_{MA}$ ,  $\hat{p}_{MA}$  als auch deren Differenzen angegeben. Auf den Seiten 27 bis 32 finden sich die grafischen Darstellungen der Analyse.

**Alter** Bei der Analyse von 17 Studien, die sowohl Daten für die Grundgesamtheit als auch für die gematchten Probanden angaben, lässt sich feststellen, dass nach dem Matchen der Altersdurchschnitt signifikant um 1,3 Jahre gestiegen ist ( $\Delta_{General-PSM} = -1,3$ ; [95%KI : -1,8; -0,8]). In Abbildung 11 auf Seite 27 sind die Altersmittelwerte der Grundgesamtheit und der gematchten Population für jede analysierte Studie dargestellt. Es lässt sich erkennen, dass die errechneten Metaanalyse-Mittelwerte vor allem durch die Studie von Magee et al. [24] beeinflusst wurde, deren Altersmittelwert mit dem mit Abstand größten Gewicht in die Analyse eingegangen ist.

**EF** Die Ejektionsfraktion ist nach Matchen mit 58,1% um 1,3 Prozentpunkte signifikant höher als vor dem Matchen (Grundgesamtheit 56,8%,  $\Delta_{General-PSM} = -1,3$ ; [95%KI : -2,4; -0,2]). Es ist allerdings fraglich, ob eine erhöhte Ejektionsfraktion von 1,3% einen Patienten merklich gesünder macht und diese Differenz somit die externe Validität negativ beeinflussen könnte.

**Geschlecht** Bezüglich des Geschlechtes lässt sich feststellen, dass nach dem Matchen anteilig signifikant weniger Männer vorhanden sind. Ihr Anteil fällt von 74,7% in der Grundgesamtheit auf 72,1% in der gematchten Gruppe ( $\Delta_{General-PSM} = 2,7$ ; [95%KI : 2,2; 3,1]).

**Diabetes** Der Anteil der Diabetiker verändert sich faktisch nicht. Der errechnete Unterschied von 0,01 Prozentpunkten ist vernachlässigbar klein ([95%KI : -1,0; 1,0]).

**Hypertonie** Nach Matchen zeigt sich im Vergleich zur Grundgesamtheit eine Erhöhung des Anteils der Hypertoniker um 1,5%. Angesichts des Konfidenzintervalls ([95%KI : -3,2; 0,2]) lassen sich jedoch keine gesicherten Schlüsse ziehen.

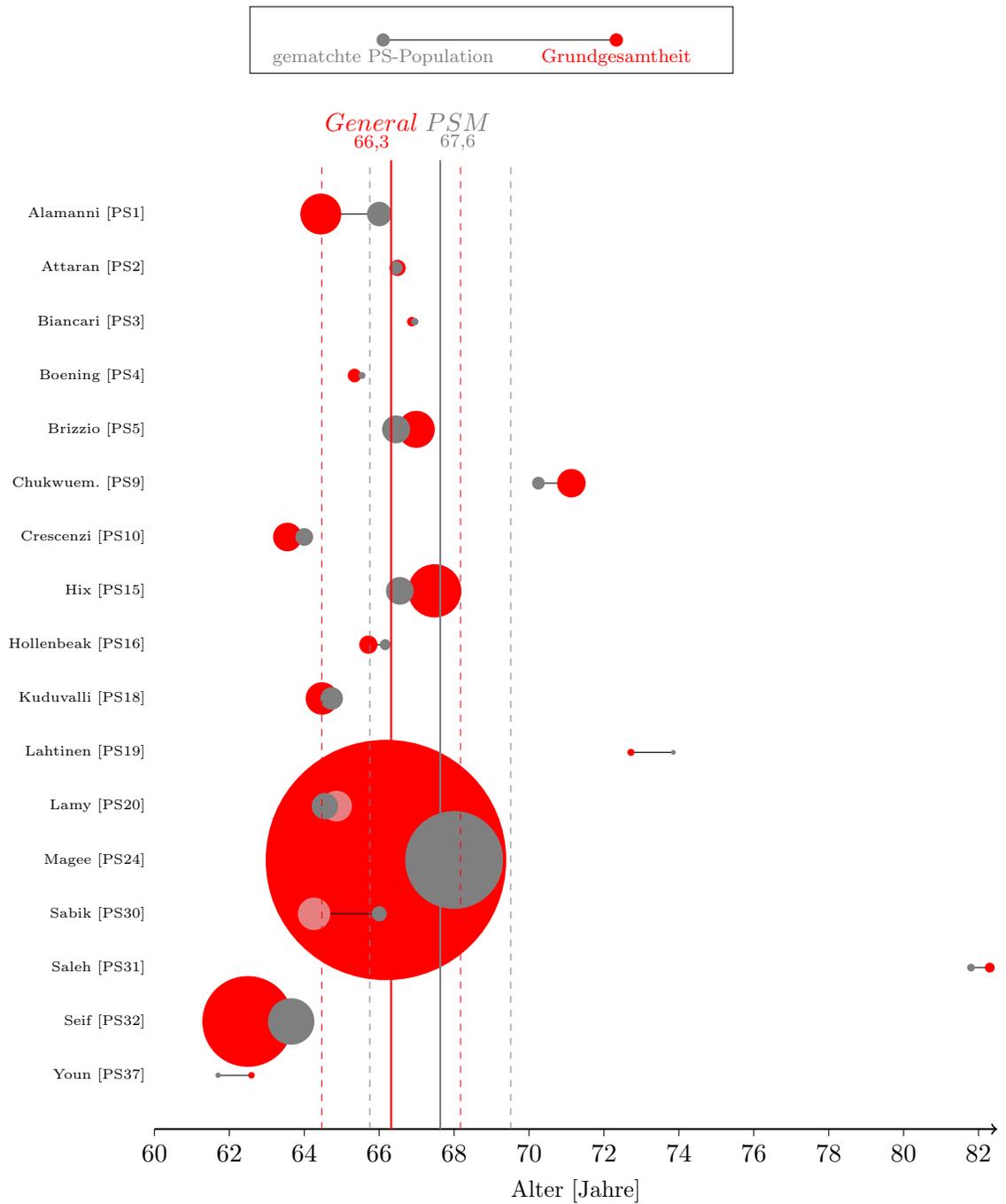
**Myokardinfarktgeschichte** In der gematchten Population gibt es mit 47,7% marginal mehr Menschen mit einer Myokardinfarktgeschichte als in der Grundgesamtheit mit 46,7%. Die Differenz von 0,7% ist jedoch fast vernachlässigbar klein und außerdem wegen des Konfidenzintervalls ([95%KI : -2,3; 0,9]) nicht generalisierbar.

Betrachtet man den durchschnittlichen Patienten vor und nach Matchen, lässt sich also beobachten, dass er signifikant älter, bezüglich seiner Ejektionsfraktion gesünder und mit geringerer Wahrscheinlichkeit männlich ist. Des Weiteren lässt sich ablesen, dass er mit höherer Wahrscheinlichkeit Hypertoniker ist und bereits einen Myokardinfarkt hatte.

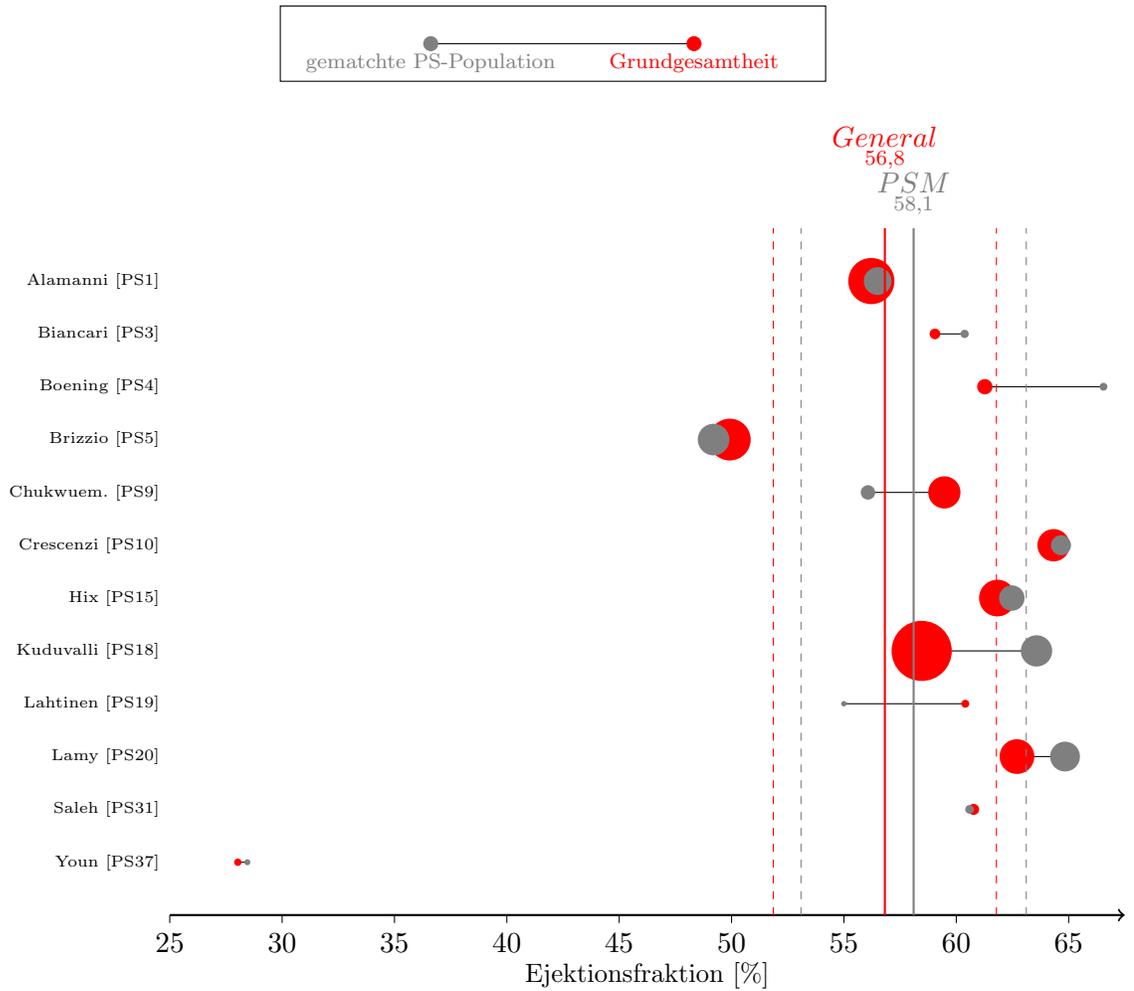
**Tabelle 4:** Within-Analyse der externen Validität. Aufgeführt sind die Meta-Mittelwerte  $\bar{x}_{MA}$  bzw. die Meta-Häufigkeiten  $\hat{p}_{MA}$  der Analyse derjenigen PS-Studien, die sowohl Angaben zu Grundgesamtheit als auch zu Probanden der gematchten Studienteilnehmer machten; zusätzlich angegeben wird die Differenz  $\Delta_{General-PSM}$  mit ihrem 95%-Konfidenzintervall und zugehörigem p-Wert.

		$n$	$\bar{x}_{MA}$	95%KI	$\Delta_{General-PSM}$	95%KI	p-Wert
$Alter$ [Jahre]	General	17	66,3	64,5 ; 68,2	-1,3	-1,8 ; -0,8	<,0001
	PSM		67,6	65,7 ; 69,5			
$EF$ [%]	General	12	56,8	51,9 ; 61,8	-1,3	-2,4 ; -0,2	0,0216
	PSM		58,1	53,1 ; 63,1			
		$n$	$\hat{p}_{MA}$	95%KI	$\Delta_{General-PSM}$	95%KI	p-Wert
$G$ [%]	General	14	74,7	70,8 ; 78,7	2,7	2,2 ; 3,1	<,0001
	PSM		72,1	68,1 ; 76,0			
$Diab.$ [%]	General	11	27,0	20,5 ; 33,6	-0,01	-1,0 ; 1,0	0,9791
	PSM		27,0	20,5 ; 33,6			
$Hyp.$ [%]	General	8	60,0	54,0 ; 66,1	-1,5	-3,2 ; 0,2	0,0778
	PSM		61,5	55,4 ; 67,6			
$MI$ [%]	General	7	46,7	27,6 ; 65,9	-0,7	-2,3 ; 0,9	0,3492
	PSM		47,4	28,2 ; 66,5			

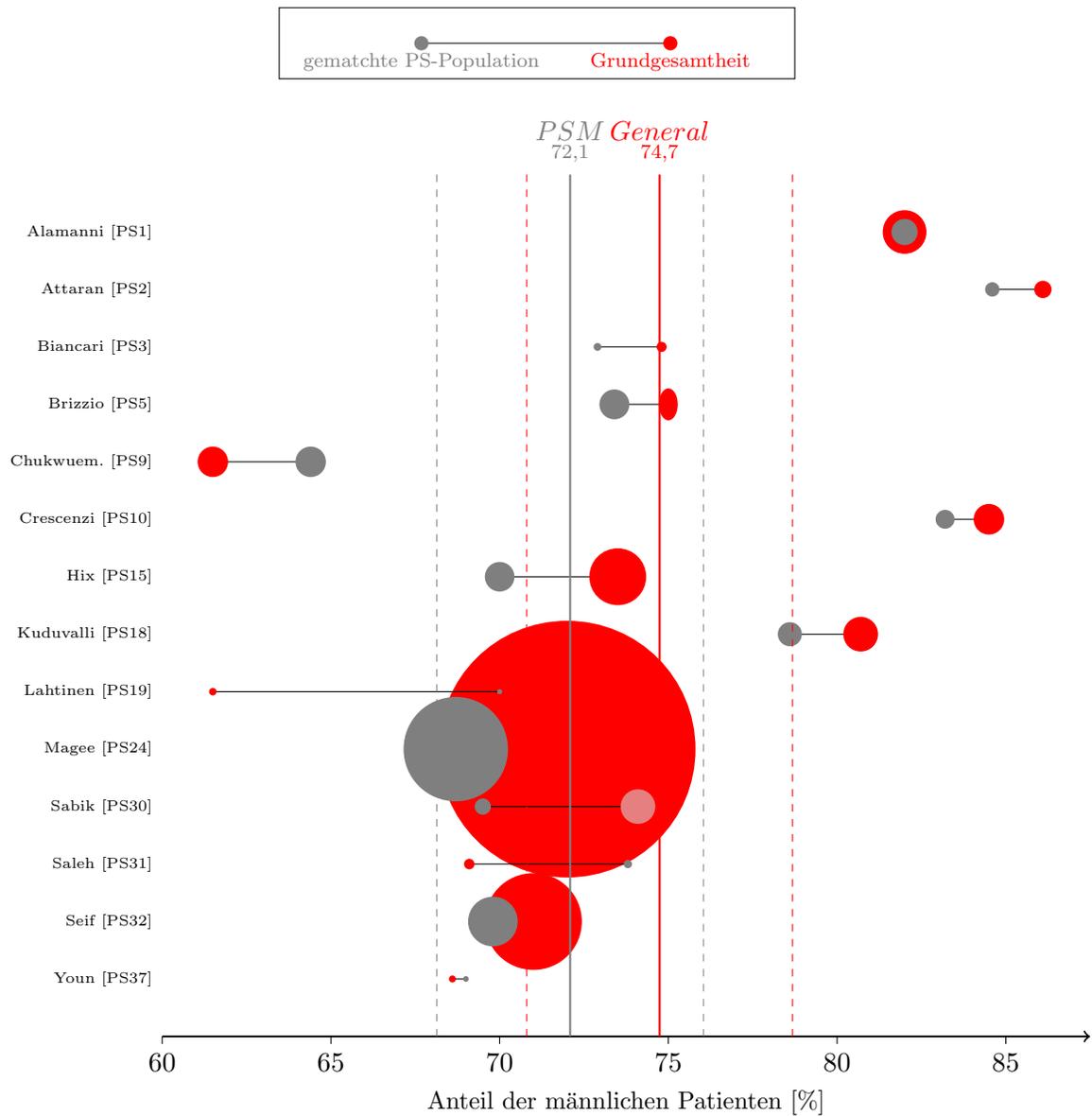
\*Geschlecht (G.), Diabetes (Diab.), Hypertonie (Hyp.)



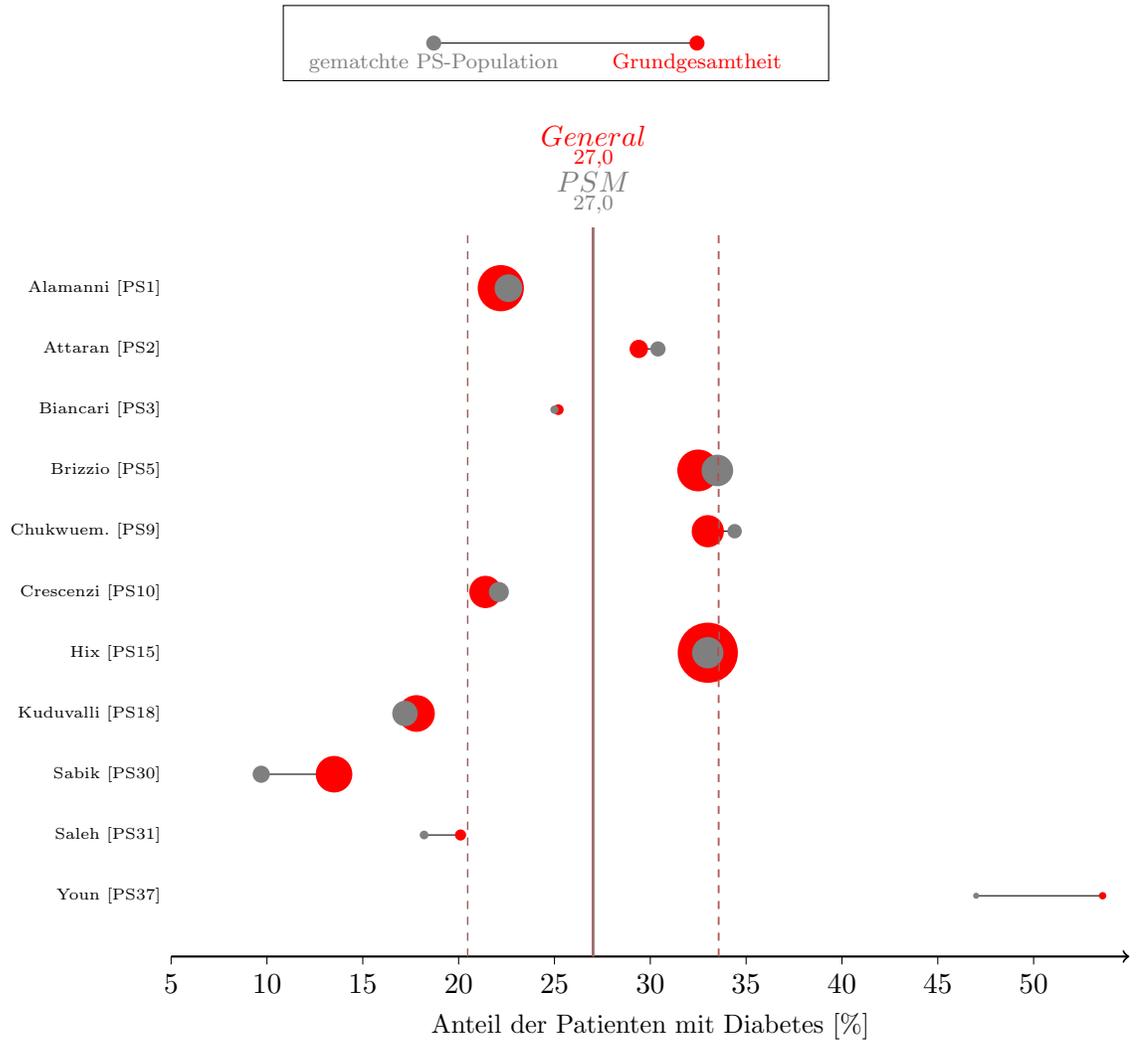
**Abbildung 11:** Within-Analyse des Alters. Jedes rot/graue Paar von Kreisen entspricht einer PS-Studie. Die Fläche der Kreise ist proportional zur Anzahl der Studienteilnehmer, die zur Gewichtung herangezogen wurde. Je größer die Fläche des Kreises, desto größer das Gewicht der entsprechenden Studie. Senkrecht dargestellt sind die Meta-Mittelwerte  $\bar{x}_{MA}$  der Within-Analyse mit Konfidenzintervall.



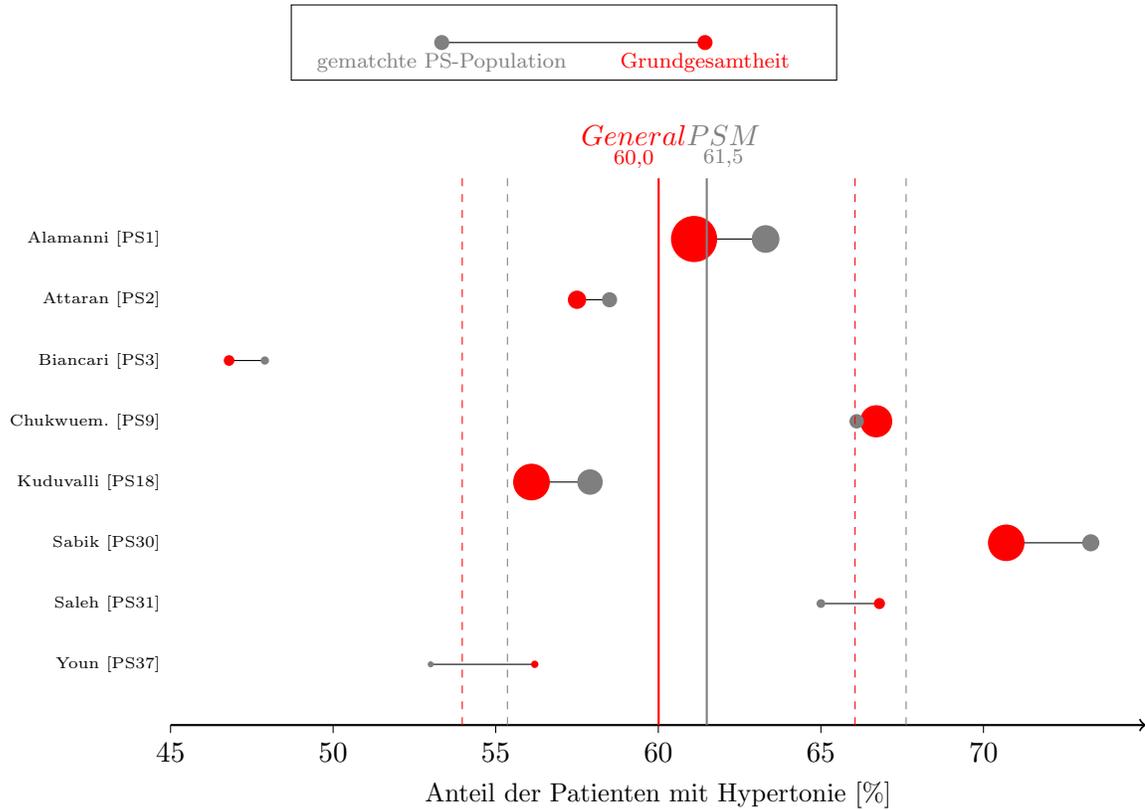
**Abbildung 12:** Within-Analyse der linksventrikulären Ejektionsfraktion. Jedes rot/graue Paar von Kreisen entspricht einer PS-Studie. Die Fläche der Kreise ist proportional zur Anzahl der Studienteilnehmer, die zur Gewichtung herangezogen wurde. Je größer die Fläche des Kreises, desto größer das Gewicht der entsprechenden Studie. Senkrecht dargestellt sind die Meta-Mittelwerte  $\bar{x}_{MA}$  der Within-Analyse mit Konfidenzintervall.



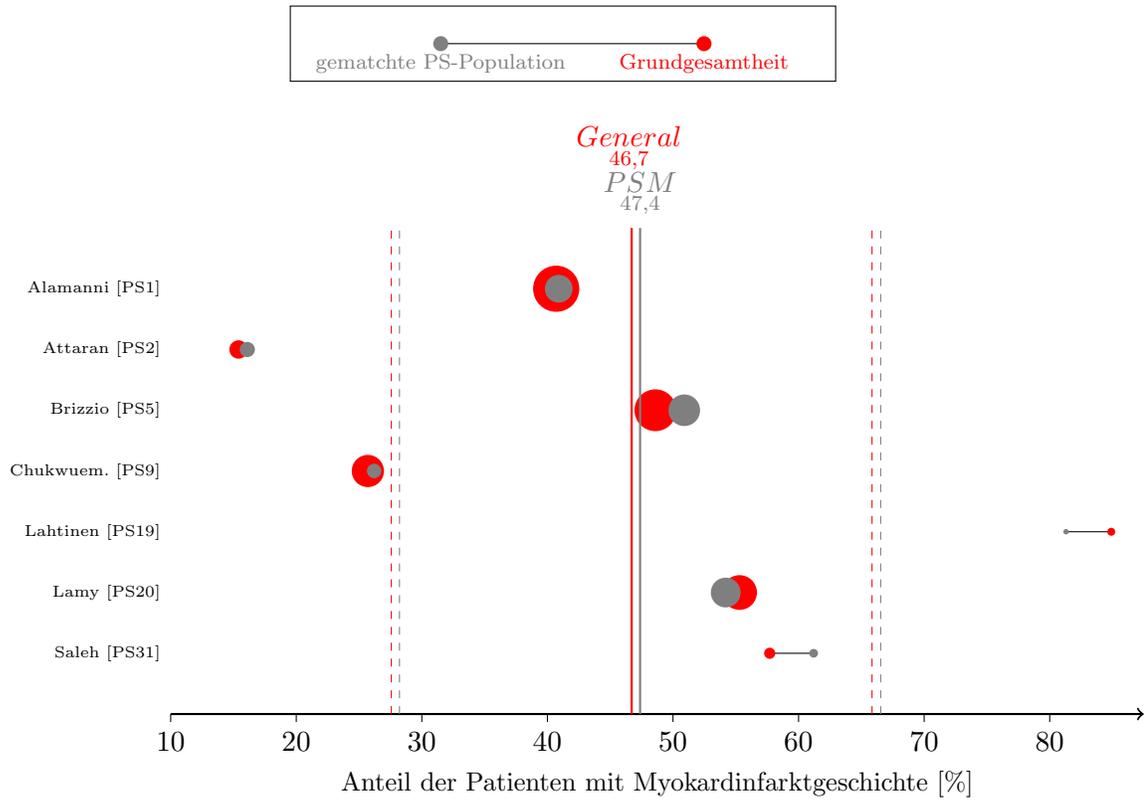
**Abbildung 13:** Within-Analyse des Anteils der männlichen Patienten. Jedes rot/graue Paar von Kreisen entspricht einer PS-Studie. Die Fläche der Kreise ist proportional zur Anzahl der Studienteilnehmer. Da im GLMM die Studien implizit gewichtet werden, entspricht hier die Größe der Kreise *nicht* exakt dem Gewicht der jeweiligen Studie. Senkrecht dargestellt sind die Meta-Häufigkeiten  $\hat{p}_{MA}$  der Within-Analyse mit Konfidenzintervall.



**Abbildung 14:** Within-Analyse des Anteils der Diabetiker. Jedes rot/graue Paar von Kreisen entspricht einer PS-Studie. Die Fläche der Kreise ist proportional zur Anzahl der Studienteilnehmer. Da im GLMM die Studien implizit gewichtet werden, entspricht hier die Größe der Kreise *nicht* exakt dem Gewicht der jeweiligen Studie. Senkrecht dargestellt sind die Meta-Häufigkeiten  $\hat{p}_{MA}$  der Within-Analyse mit Konfidenzintervall.



**Abbildung 15:** Within-Analyse des Anteils der Hypertoniker. Jedes rot/graue Paar von Kreisen entspricht einer PS-Studie. Die Fläche der Kreise ist proportional zur Anzahl der Studienteilnehmer. Da im GLMM die Studien implizit gewichtet werden, entspricht hier die Größe der Kreise *nicht* exakt dem Gewicht der jeweiligen Studie. Senkrecht dargestellt sind die Meta-Häufigkeiten  $\hat{p}_{MA}$  der Within-Analyse mit Konfidenzintervall.



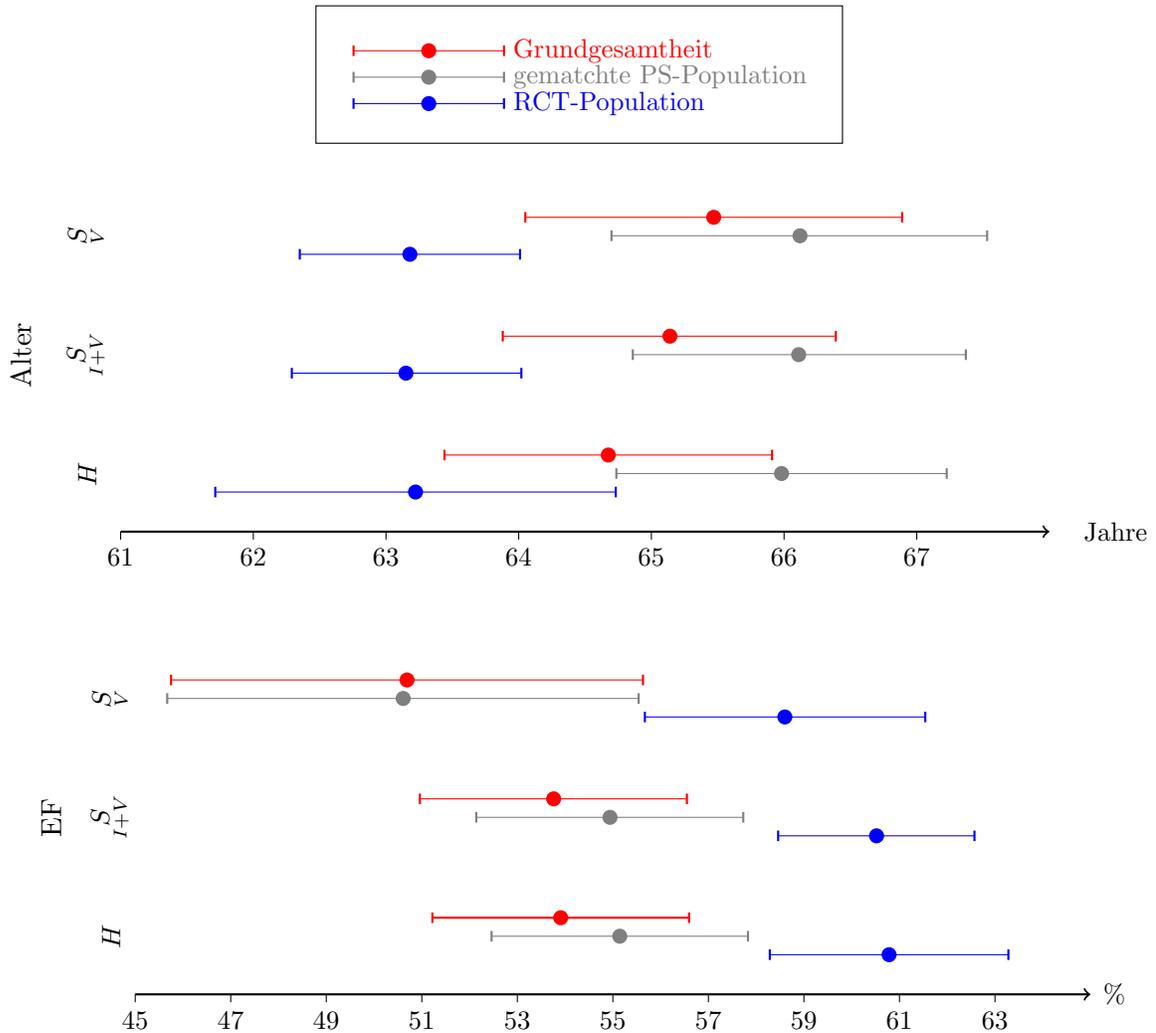
**Abbildung 16:** Within-Analyse des Anteils der Patienten mit Myokardinfarktgeschichte. Jedes rot/graue Paar von Kreisen entspricht einer PS-Studie. Die Fläche der Kreise ist proportional zur Anzahl der Studienteilnehmer. Da im GLMM die Studien implizit gewichtet werden, entspricht hier die Größe der Kreise *nicht* exakt dem Gewicht der jeweiligen Studie. Senkrecht dargestellt sind die Meta-Häufigkeiten  $\hat{p}_{MA}$  der Within-Analyse mit Konfidenzintervall.

### 3.4 Sensitivitätsanalyse

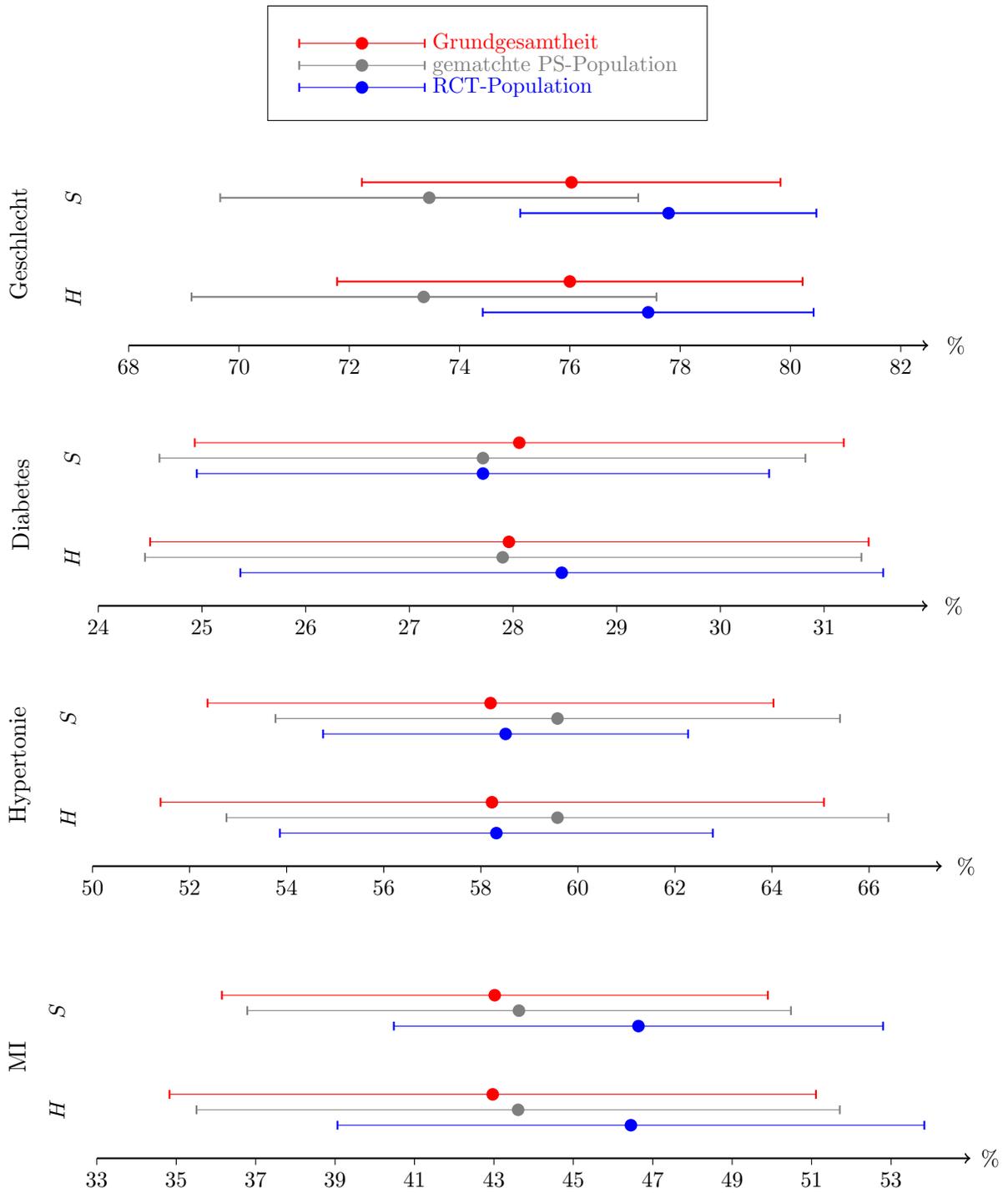
Ziel der Sensitivitätsanalyse war es herauszufinden, ob ein einfacheres Analyse-Modell, welches bestimmte Eigenschaften der Studien dieses systematischen Reviews vernachlässigt (wie etwa die Korrelation von Teilnehmern in gemachten PS-Studien und der Grundgesamtheit, siehe Abschnitt 2.3.3), andere Ergebnisse errechnet, als das bestmögliche Modell in der Hauptanalyse. In der Sensitivitätsanalyse wurde das Random-Effects-Modell zu Grunde gelegt. Des Weiteren wurde nicht mit der Fallzahl, sondern mit der inversen Varianz gewichtet. Eine Übersicht über die Unterschiede zwischen Haupt- und Sensitivitätsanalyse bietet Tabelle 1 auf Seite 11. Da bei der Gewichtung mit inverser Varianz ebenfalls beachtet werden musste, dass bei den metrischen Merkmalen Alter und EF oft die Standardabweichung imputiert wurde, finden sich im Folgenden für die metrischen Merkmale zwei Sensitivitätsanalysen: eine mit allen verfügbaren (also auch den imputierten) Daten, und eine mit ausschließlich in den Studien explizit angegebenen Standardabweichungen. Aus der Abbildung 4 auf Seite 19 geht hervor, in welchem Maße bei welchen Merkmalen imputiert wurde. Es ist deutlich erkennbar, dass insbesondere bei der linksventrikulären Ejektionsfraktion viele Daten imputiert wurden, da die gesuchten Lage- und Streuungsmaße ursprünglich nicht gegeben waren. Eine vollständige Auflistung der Ergebnisse der Sensitivitätsanalyse findet sich in den Tabellen 2 und 3 auf Seite 23 und 24 und sind den Ergebnissen der Hauptanalyse gegenübergestellt.

Dieses Kapitel beschränkt sich auf die grafische Darstellung der Ergebnisse. In Abbildung 17 bzw. 18 finden sich die errechneten Mittelwerte  $\bar{x}_{MA}$  bzw.  $\hat{p}_{MA}$  für jedes Merkmal, Studienpopulation und Art der Analyse gegenübergestellt. Nach Betrachtung der Grafiken und Einbeziehung der entsprechenden Tabellen lässt sich feststellen, dass die errechneten Mittelwerte bzw. Häufigkeiten in der Sensitivitätsanalyse zwar teilweise von denen der Hauptanalyse abweichen, jedoch die Unterschiede zwischen Haupt- und Sensitivitätsanalyse bedeutend kleiner sind, als die Unterschiede zwischen den verschiedenen Studienpopulationen. Mag sich auch das ein oder andere Ergebnis der Sensitivitätsanalyse um wenige Prozentpunkte von seinem Gegenstück der Hauptanalyse unterscheiden, bleibt die Tendenz erhalten. Für die Merkmale Geschlecht, Hypertonie und Myokardinfarktgeschichte sind die Unterschiede vernachlässigbar klein. Beim Anteil der Diabetiker scheint sich lediglich der Wert für die RCTs so zu verschieben, dass eine leicht abweichende Anordnung der Werte entsteht. Die grobe Anordnung der Meta-Mittelwerte des Alters zwischen den Studienpopulationen bleibt weitgehend bestehen. Nur bei linksventrikulärer Ejektionsfraktion ändert sich das Bild deutlich. Dies wird verständlich, wenn man in Grafik 4 auf Seite 19 betrachtet, wie viele Daten bei der Ejektionsfraktion imputiert werden mussten und deshalb bei der Sensitivitätsanalyse der ausschließlich vollständigen Daten ausgeschlossen wurden. Durch die starke Dezimierung der analysierten Studien wächst das Konfidenzintervall und die Mittelwerte werden verschoben. Zu den Konfidenzintervallen lässt sich im Allgemeinen feststellen, dass diese in der Hauptanalyse stets größer sind als in der Sensitivitätsanalyse. Die Hauptanalyse schätzt ihre Ergebnisse also konservativer.

Obwohl die Sensitivitätsanalyse einige wichtige Annahmen vernachlässigt, liefert sie zwar leicht, aber nicht grundsätzlich von der Hauptanalyse unterschiedliche Ergebnisse. Der Einfluss der Studienpopulation ist weitaus größer als der der angewandten Analyse.



**Abbildung 17:** Vergleichende Darstellung der ermittelten Meta-Mittelwerte  $\bar{x}_{MA}$  mit ihren 95%- Konfidenzintervallen für Alter und linksventrikuläre Ejektionsfraktion aller drei Studienpopulationen in Haupt-, sowie Sensitivitätsanalyse. Hauptanalyse (H), Sensitivitätsanalyse (S), imputierte *und* vollständige Daten (I+V), ausschließlich vollständige Daten (V).



**Abbildung 18:** Vergleichende Darstellung der ermittelten Meta-Häufigkeiten  $\hat{p}_{MA}$  mit ihren 95%-Konfidenzintervallen für Geschlecht, Diabetes, Hypertonie und Myokardinfarktgeschichte aller drei Studienpopulationen in Haupt-, sowie Sensitivitätsanalyse. Hauptanalyse (H), Sensitivitätsanalyse (S).

## 4 Diskussion

Ziel des systematischen Reviews war es, randomisiert-kontrollierte Studien und gematchte Propensity-Score-Studien hinsichtlich ihrer externen Validität zu untersuchen. Dazu wurden für die Merkmale Alter, linksventrikuläre Ejektionsfraktion, Geschlecht, Diabetes, Hypertonie und Myokardinfarktgeschichte die Meta-Mittelwerte  $\bar{x}_{MA}$  bzw. die Meta-Häufigkeiten  $\hat{p}_{MA}$  in den drei Populationen (Grundgesamtheit, gematchte PS- und RCT-Studienpopulation) berechnet und die Differenzen ermittelt (Between-Analyse). Um festzustellen, wie sich die durchschnittlichen Merkmale der Probanden durch Matchen verändern, wurden in der Within-Analyse  $\bar{x}_{MA}$  bzw.  $\hat{p}_{MA}$  ausschließlich derjenigen PS-Studien ermittelt, die Angaben sowohl zur Grundgesamtheit, als auch zu der gematchten PS-Studienpopulation enthielten. Die Differenz zwischen beiden Populationen gibt Aufschluss über den Einfluss des Matchens auf die durchschnittlichen Eigenschaften der Probanden.

### Einschränkungen

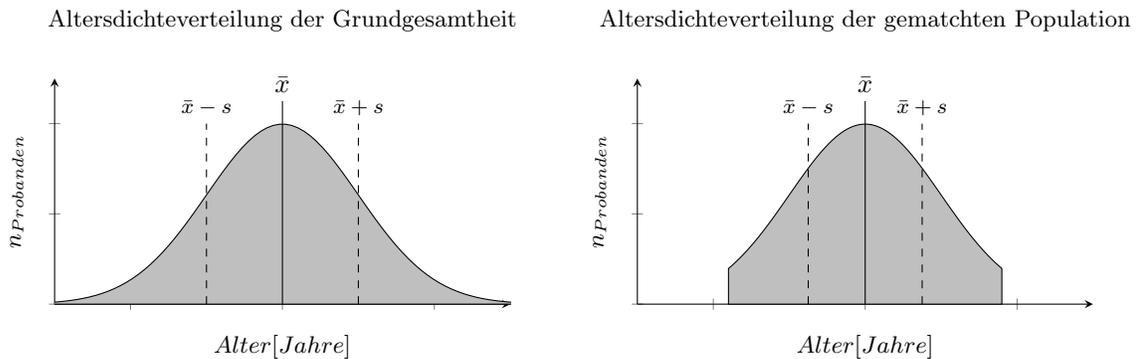
Bei der Bewertung der Ergebnisse sind jedoch Einschränkungen vorzunehmen. Als Vergleichspunkt, an dem die externe Validität von RCTs und gematchten PS-Studien überprüft wurde, diente in diesem systematischen Review die Grundgesamtheiten, die in den eingeschlossenen Propensity-Score-Studien angegeben wurden. Es wird davon ausgegangen, dass diese die durchschnittlichen Patienten einer On- bzw. Off-Pump Bypass-Operation repräsentieren. Es ist jedoch wichtig festzustellen, dass hier nur solche Patienten in die Analyse eingehen, die gematchten PS-Studien zu Grunde lagen. Andere Kollektive, wie z.B. die Grundgesamtheiten von PS-Studien, die mit Stratifizierung, Regression oder Gewichtung mit inverser Behandlungsmöglichkeit arbeiten und Patienten von Zentren, an denen RCTs durchgeführt werden, bleiben unbetrachtet. Dies könnte zu einem Problem für die Belastbarkeit der hier getroffenen Aussagen zur externen Validität führen, wenn Zentren, an denen gematchte Propensity-Score-Studien durchgeführt werden, ein anderes Patientengut aufweisen als Zentren, an denen beispielsweise RCTs durchgeführt werden. In einer Studie von McKee et al. [1] wird festgestellt, dass RCTs häufiger an universitären Zentren durchgeführt werden, während nicht-randomisierte Studien oft nicht-universitäre Zentren einschließen. Im Allgemeinen werden auf Grund von höherer Expertise und technischer Ausstattung kritische Patienten eher in universitäre Zentren verlegt. Es ist nicht auszuschließen, dass die hier verwendete Grundgesamtheit gar nicht den durchschnittlichen Patienten repräsentiert, sondern eventuell sogar kränker ist. In diesem Falle würde der Vergleich zwischen der Grundgesamtheit und der RCT-Studienpopulation bzw. der gematchten PS-Studienpopulation keine belastbaren Ergebnisse produzieren. Die Tatsache, dass die gematchten Probanden aus dieser Grundgesamtheit hervorgehen, also nicht statistisch unabhängig sind, wird im Modell der Hauptanalyse berücksichtigt und hat keinen einschränkenden Einfluss.

Beim Auslesen der metrischen Daten aus den verschiedenen Studien stellte sich oft das Problem, dass die zur späteren Analyse nötigen einheitlichen Lage- und Streuungsmaße nicht gegeben waren und diese mit Hilfe anderer Daten und Annahmen imputiert werden mussten. Nur für die Ermittlung der Standardabweichungen, die für die Sensitivitätsanalyse benötigt wurden, wurde auf die Imputation von Daten aus der Hallenser 3CAB-Studie [27] zurückgegriffen. Da diese Imputation recht grob ist, wurde in der Sensitivitätsanalyse eine vergleichende Analyse mit imputierten und vollständigen Daten, sowie auch unter Ausschluss der imputierten Daten durchgeführt. Diese recht grobe Form der Imputation spielte keine Rolle in der Hauptanalyse. Doch wurden auch bei der Ermittlung des Mittelwertes der metrischen Merkmale oftmals Daten unter der Annahme, die Daten seien gleichverteilt, aus Angaben wie „60% der Patienten hatten eine EF von  $< 50\%$ “ angenähert bzw. die Mediane mit Mittelwerten gleichgesetzt. Sollten diese Annahmen fehlerhaft sein, ist es möglich, dass Mittelwerte verfälscht und dadurch die Ergebnisse bezüglich einer Aussage zur externen Validität nicht belastbar sind.

Ein wichtiger potentieller Störfaktor dieses systematischen Reviews ist die fragliche Objektivität der Merkmale und ihre Messung. Während sich das Alter und das Geschlecht eines Probanden meist zweifelsfrei und eindeutig bestimmen lassen, können z.B. für Diabetes und Hypertonie unterschiedliche Diagnosekriterien zu Grunde gelegt werden. Auch die Messung der linksventrikulären Ejektionsfraktion kann sich je nach Untersucher und Untersuchungsgerät um wenige Prozentpunkte unterscheiden. Besonders deutlich wurde diese Problematik bei der Datenerhebung der Myokardinfarktgeschichte. Der Begriff „Myokardinfarktgeschichte“ bezog sich auf sehr unterschiedliche Zeiträume; schwankend zwischen „in den letzten 24 Stunden“ bis zu jedem beliebigen Zeitpunkt im Leben bis dato. Im vorliegenden systematischen Review wurden nur solche Daten aufgenommen, die keinerlei Einschränkungen zum Zeitpunkt des Myokardinfarkts machten. Es konnten deshalb nur rund 50% der verfügbaren Daten in dieses systematische Review eingehen, worunter die Aussagekraft der gewonnenen Ergebnisse litt. Ein weiterer Unsicherheitsfaktor ist, dass z.B. das Vorliegen eines Diabetes oder einer Hypertonie auch eine vom Patienten angegebene Information sein kann, die der Untersucher abfragt, aber nicht nachprüft. Die Qualität der erhobenen Daten würde in diesem Fall darunter leiden, dass Patienten oftmals die Fülle ihrer Erkrankungen nicht mehr überblicken oder Informationen auf Grund mangelnden Fachwissens durcheinanderbringen. Je nach Art der Datenerhebung und zu Grunde gelegten Diagnosekriterien könnten sich gefundene Differenzen zwischen Studienpopulationen also nicht nur auf tatsächliche Unterschiede zwischen den Populationen zurückführen lassen, sondern auch auf strukturelle Unterschiede in der Studiendurchführung.

Bei der Analyse der metrischen Daten der gematchten Propensity-Score-Studien sollte herausgefunden werden, inwieweit nicht gematchte Probanden die externe Validität durch eine Veränderung der durchschnittlichen Eigenschaften der Population verschlechtern. Hierbei wurde

der für die gematchten PS-Studienpopulation errechnete Mittelwert mit dem der Grundgesamtheit verglichen. Wie in Abbildung 19 dargestellt, könnte sich hierbei jedoch folgendes Problem ergeben: Zu Anschauungszwecken geht man beim dargestellten Beispiel von einer Normalverteilung des Alters und davon aus, dass insbesondere Probanden am unteren und oberen Ende des Spektrums nicht gematcht werden können. Fallen nun am oberen und unteren Ende gleich viele Probanden heraus, lassen sich die Ergebnisse der Studie für die herausgefallenen Probanden nicht mehr generalisieren und die Studie verliert an externer Validität, obwohl der Mittelwert gleich bleibt.



**Abbildung 19:** Dargestellt wird beispielhaft die Anzahl der Probanden nach Alter sowie der Mittelwert  $\bar{x}$  und die Standardabweichung  $s$  der Verteilung. Fallen gleichmäßig Probanden mit besonders hohem oder niedrigem Alter heraus, bleibt der Mittelwert  $\bar{x}$  gleich, die Standardabweichung  $s$  wird jedoch kleiner.

Einzig die Standardabweichung würde kleiner. Die Standardabweichung lässt sich jedoch nicht als Maß für eine Einschränkung des Studienkollektivs gebrauchen, da sie von der Fallzahl  $n$  abhängig ist (siehe Formel (21)). Bei schrumpfender Fallzahl würde die Standardabweichung größer.

$$s = \sqrt{\frac{1}{n-1} * \sum_{i=1}^n (x_i - \bar{x})^2} \quad (21)$$

## Interpretation der gewonnenen Ergebnisse

**Randomisiert-kontrollierte Studien** Nachdem die Beschränkungen der Arbeit aufgezeigt wurden, sollen die Ergebnisse kritisch interpretiert werden. Betrachtet man zunächst die externe Validität von randomisiert-kontrollierten Studien, so lässt sich hervorheben, dass Teilnehmer an randomisiert-kontrollierten Studien im Schnitt eine deutlich und signifikant höhere linksventrikuläre Ejektionsfraktion aufweisen (60,8%, [95%KI : 58,3; 63,3]) als die Grundgesamtheit (53,9%, [95%KI : 51,2; 56,6]) und als die Teilnehmer der gematchten PS-Studien (55,1%, [95%KI : 52,5; 57,8]). Der Normbereich der linksventrikulären Ejektionsfraktion wird

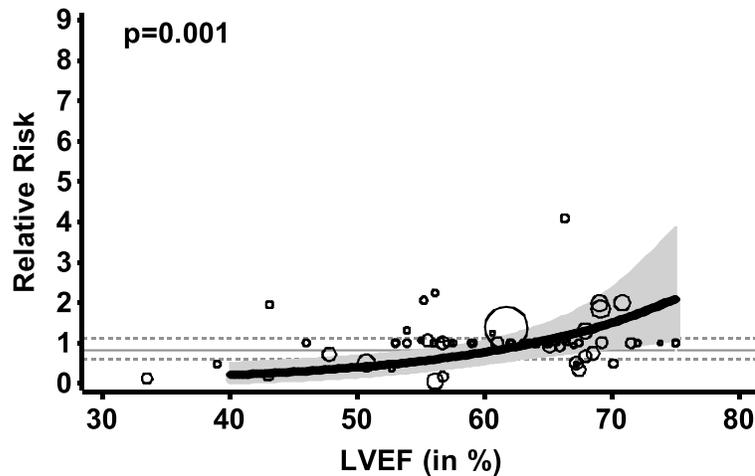
**Tabelle 5:** Das relative Mortalitätsrisiko (RR) für Off-Pump-Bypass-Operationen gegenüber einem On-Pump-Verfahren in Abhängigkeit von der linksventrikulären Ejektionsfraktion (LVEF) [28].

LVEF	RR	95%KI
50%	0,40	0,23 ; 0,69
55%	0,56	0,37 ; 0,83
60%	0,77	0,56 ; 1,06

in der Literatur mit  $> 55\%$  angegeben [13, S. 210]. Es wird deutlich, dass die Grundgesamtheit eine leicht verminderte Ejektionsfraktion aufweist, während die RCT-Studienpopulation eine normale Ejektionsfraktion hat.

In einem systematischen Review von Kuß und Börgermann [28] wurde der Einfluss des Alters, des Anteils von Frauen und der linksventrikulären Ejektionsfraktion auf die Überlegenheit von Off-Pump- gegenüber On-Pump-Bypass-Operationen in Hinblick auf die Zielgrößen Mortalität, Myokardinfarkt, Schlaganfall und Kammerflimmern untersucht. Dabei wurde für die Zielgröße Mortalität festgestellt, dass das relative Risiko der Offpump-Operation mit steigender Ejektionsfraktion steigt. Bei niedriger linksventrikulärer Ejektionsfraktion ist die Off-Pump-Methode bezüglich der Sterblichkeit überlegen, bei hoher linksventrikulärer Ejektionsfraktion ist die On-Pump-Methode überlegen. Die durchschnittliche Ejektionsfraktion der RCT-Studienpopulation beträgt im vorliegenden systematischen Review  $60,8\%$  ([95%KI : 58,3; 63,3]) und in der Grundgesamtheit  $53,9\%$  ([95%KI : 51,2; 56,6]). In Tabelle 5 sowie in Abbildung 20 auf Seite 40 sind die entsprechenden, von Kuß und Börgermann errechneten relativen Risiken für Mortalität in Abhängigkeit der Ejektionsfraktion aufgelistet bzw. graphisch dargestellt. Es wird deutlich, dass die externe Validität von RCTs bezüglich der Ejektionsfraktion deutlich eingeschränkt ist, weil die Überlegenheit von Off- bzw. On-Pump-Verfahren stark von der Ejektionsfraktion abhängig ist. Der Effekt von Off-Pump-Bypass-Operationen wird also in den im vorliegenden systematischen Review untersuchten RCTs *unterschätzt*. Die Masse der RCTs, die anscheinend bezüglich der Ejektionsfraktion regelmäßig und systematisch kränkere Patienten nicht in das Studienkollektiv aufnehmen, birgt ein Risiko: Die externe Validität ist eingeschränkt, denn die Ergebnisse der Studien sind nicht für alle Patienten gültig und Patienten mit niedriger Ejektionsfraktion werden vernachlässigt. Für diese Patienten gibt es keine für sie gültigen Studienergebnisse und somit auch keine Erkenntnisse darüber, welche Behandlung für sie optimal wäre.

Die Teilnehmer der RCTs sind durchschnittlich 1,5 Jahre jünger als die Grundgesamtheit. Diese Differenz ist zwar nicht signifikant, zeigt jedoch einen Trend an und ließe sich ebenfalls mit einer Einengung des Kollektivs auf jüngere Patienten erklären, da ältere Patienten unter anderem öfter komorbide sind und so mit den in der Einleitung aufgeführten medizinischen und wissenschaftlichen Motivationen ausgeschlossen werden. In der Studie von Kuß und Börgermann



**Abbildung 20:** In dem Streudiagramm, das ebenfalls aus Kuß und Börgermann [28] entnommen wurde, ist der Zusammenhang zwischen linksventrikulärer Ejektionsfraktion und relativem Mortalitätsrisiko für Off-Pump-Bypass-Operationen gegenüber einem On-Pump-Verfahren dargestellt. Jeder Kreis repräsentiert eine Studie. Die Fläche der Kreise ist proportional zur Anzahl der Studienteilnehmer. Die schwarze Linie gibt das relative Mortalitätsrisiko für Off-Pump-Bypass-Operationen gegenüber dem On-Pump-Verfahren in Abhängigkeit der Ejektionsfraktion an. Linksventrikuläre Ejektionsfraktion (LVEF).

wurde kein Einfluss des Alters auf das relative Risiko von Off-Pump- gegenüber On-Pump-Bypass-Operationen in Hinblick auf die Zielgrößen Mortalität, Myokardinfarkt, Schlaganfall und Kammerflimmern festgestellt. Dementsprechend kann hier davon ausgegangen werden, dass ein Altersunterschied von 1,5 Jahren die externe Validität nicht negativ beeinflusst.

Die Teilnehmer an RCTs haben eine 1,4% höhere Wahrscheinlichkeit als die Grundgesamtheit, männlich zu sein. Es ist bekannt, dass Frauen zwar weniger häufig einen Myokardinfarkt erleiden (Männer erkranken 3,4-mal so häufig wie Frauen), jedoch versterben Frauen häufiger und entwickeln öfter schwere Komplikationen wie Herzinsuffizienz oder Schlaganfälle. Die Ursachen hierfür sind noch nicht gefunden [14]. Ungeachtet dieser schlechteren Prognose konnte die Studie von Kuß und Börgermann [28] keinen Einfluss der Geschlechterverteilung auf das relative Risiko von Off-Pump- gegenüber On-Pump-Bypass-Operationen feststellen. Die externe Validität wird durch den gering erhöhten Anteil von Männern nicht negativ beeinflusst. Bezüglich des Anteils der Diabetiker und der Hypertoniker ließen sich nur marginale Unterschiede zur Grundgesamtheit feststellen. Der Anteil der Patienten mit Myokardinfarktgeschichte scheint bei RCTs erhöht (um 3,5% höher). Da jedoch die oben bereits diskutierten Einschränkungen bei der Erhebung der Myokardinfarktgeschichte in einem sehr großen Konfidenzintervall ( $[95\%KI : -14,5; 7,5]$ ) für diese Differenz resultierten, ist der gefundene Unterschied zwischen Grundgesamtheit und RCTs schwierig abschließend zu bewerten und nicht belastbar.

Die Ergebnisse dieser Arbeit unterstützen die Erkenntnisse von Hordijk-Trion et al. [3]. Dort wurde festgestellt, dass bei der Erstellung von Studien zu koronaren Interventionen eine rigorose Auswahl der Patienten vorgenommen wurde und dass die Klinikpopulation im Schnitt älter und komorbider, also weniger gesund war, als die RCT-Studienpopulation. In der vorliegenden Arbeit wurde zwar nur eine klinisch relevante Einschränkung der externen Validität bezüglich der linksventrikulären Ejektionsfraktion festgestellt, doch auch andere Merkmale wichen in ihrer Ausprägung von der Grundgesamtheit ab. Wie bereits in der Einführung beschrieben gibt es nur wenige Studien, die RCTs eine einwandfreie externe Validität zusprechen. Eine der wenigen Studien ist die von Sorg et al. [7]; dort wurden zwar keine groben, klinisch relevanten Unterschiede zwischen RCT-Studienpopulation und Klinikpopulation gefunden, doch gab es dort, wie auch im vorliegenden systematischen Review, für die weiteren untersuchten Merkmale mehrere kleine Kritikpunkte. Insgesamt befinden sich die hier gewonnenen Ergebnisse in Einklang mit der Studienlage zur externen Validität von RCTs.

**Gematchte Propensity-Score-Studien** Interpretiert man die Ergebnisse der Between-Analyse für die gematchten Propensity-Score-Studien, so zeigen sich in drei von sechs Merkmalen signifikante Abweichungen zur Grundgesamtheit. Die Patienten waren 1,3 Jahre älter, hatten eine um 1,2 Prozentpunkte höhere linksventrikuläre Ejektionsfraktion und es gab 2,6 Prozentpunkte weniger Männer als in der Grundgesamtheit. In Anbetracht der Erkenntnisse des systematischen Reviews von Kuß und Börgermann [28], sind die hier gefundenen Differenzen zwischen gematchter PS-Studienpopulation und Grundgesamtheit nicht als klinisch relevant zu bewerten. Es ist davon auszugehen, dass eine um 1,2 Prozentpunkte erhöhte linksventrikuläre Ejektionsfraktion die externe Validität nicht negativ beeinflusst. Bezüglich des Anteils der Diabetiker, Hypertoniker oder Patienten mit Myokardinfarktgeschichte lassen sich keine signifikanten Unterschiede feststellen. Die errechneten Differenzen sind vernachlässigbar klein; selbst im Falle des um 1,4% ([95%KI : -3,1; 0,4]) erhöhten Anteils von Hypertonikern ist aus klinischer Sicht nicht davon auszugehen, dass dieser Unterschied zwischen Grundgesamtheit und gematchter PS-Studienpopulation die externe Validität der gematchten PS-Studien herabsetzt.

**Within-Analyse** In der Within-Analyse sollte festgestellt werden, wie sich die durchschnittlichen Merkmale der Teilnehmer an gematchten PS-Studien nach Matches verändert haben. Der durchschnittliche Teilnehmer wurde 1,3 Jahre älter und hatte eine 1,3% höhere Ejektionsfraktion. Des Weiteren war er mit einer 2,7% niedrigeren Wahrscheinlichkeit männlich. Der Anteil der Frauen war also nach Matches erhöht. Der Anteil der Diabetiker und derjenigen mit Myokardinfarktgeschichte veränderte sich beinahe nicht. Der Anteil der Hypertoniker sank um 1,5%, wodurch die externe Validität aber nicht beeinträchtigt wird. Es lässt sich zusammenfassend sagen, dass vor allem Männer, junge Patienten und solche mit einer besonders niedrigen Ejektionsfraktion offensichtlich durch fehlende Partner beim Matchen nicht mit in die Studi-

en aufgenommen wurden. Die externe Validität wird jedoch nicht in klinisch relevanter Weise negativ beeinflusst. Da die in der Within-Analyse ermittelten Veränderungen durch Matching in etwa den in der Between-Analyse errechneten Differenzen zwischen Grundgesamtheit und PS-Studienpopulation entsprechen, ist davon auszugehen, dass die in der Between-Analyse gefundenen Differenzen zwischen Grundgesamtheit und PS-Studienpopulation tatsächlich durch Matching bedingt sind und nicht durch zentrumsbedingtes unterschiedliches Patientengut.

**Sensitivitätsanalyse** Zuletzt sollen die Ergebnisse der Sensitivitätsanalyse in Kontrast zur Hauptanalyse gesetzt werden. Ziel der Sensitivitätsanalyse war es, herauszufinden, inwieweit die gefundenen Unterschiede in den Merkmalsausprägungen durch die Studienpopulation selbst determiniert sind, und welchen Anteil die statistische Modellierung und die Grundannahmen am Ergebnis haben. Es stellte sich heraus, dass auch mit einer einfacheren statistischen Modellierung, die wichtige Annahmen zur Datenlage außer Acht ließ, ungefähr die gleichen Ergebnisse ermittelt wurde. Die Anordnung der drei Populationen änderte sich nur marginal. Die Hauptanalyse ermittelte Schätzer mit einem meist größeren Konfidenzintervall als die Sensitivitätsanalyse, sie ist also konservativer. Einzig bei der Analyse der Ejektionsfraktion ließ sich feststellen, dass bei der Sensitivitätsanalyse unter Ausschluss der imputierten Daten eine deutlich niedrigere durchschnittliche Ejektionsfraktion berechnet wurde. Die Imputation von Standardabweichungen mit Hilfe von Ranges aus der 3CAB-Studie [27] und Annahmen von Hozo et al. [22] verfälschte die Ergebnisse; es wurden zu hohe Ejektionsfraktionen imputiert. Da aber nur in der Sensitivitätsanalyse auf die Imputation von Ranges zurückgegriffen wurde, spielen diese Erkenntnisse für die Ergebnisse der Hauptanalyse keine Rolle.

**Fazit** Zusammenfassend muss festgestellt werden, dass die externe Validität von randomisiert-kontrollierten Studien eingeschränkt ist. Klinisch betrachtet ist die externe Validität von RCTs auf Grund der deutlich erhöhten linksventrikulären Ejektionsfraktion in dieser Population deutlich beeinträchtigt. Es kann davon ausgegangen werden, dass die weiteren Differenzen keinen erheblichen negativen Einfluss auf die externe Validität besitzen. Die Ergebnisse sind im Einklang mit der Großzahl bisher veröffentlichter Studien zur externen Validität von RCTs, in denen immer wieder eine eingeschränkte externe Validität festgestellt wird. Die externe Validität von gematchten PS-Studien zeigte sich nicht in klinisch relevantem Maße eingeschränkt. Die RCTs sind den gematchten PS-Studien zwar bezüglich ihrer externen Validität unterlegen, doch balancieren nur sie für bekannte *und* unbekannte Confounder. Nur RCTs stellen eine Strukturgleichheit der Therapiegruppen sicher. Gematchte Propensity-Score-Studien sind jedoch in bereits gut erforschten Fachgebieten wie z.B. der Kardiochirurgie eine billigere, einfachere und extern validere Alternative zu RCTs, da es dort unwahrscheinlich ist, dass es unbekannte Confounder gibt, die von den bereits bekannten und einbezogenen Confoundern völlig unabhängig sind.

## 5 Zusammenfassung

Bei der Prüfung von Therapien in vergleichenden Studien kann auf verschiedene Verfahren zur Bildung von Behandlungsgruppen zurückgegriffen werden, um eine möglichst hohe interne und externe Validität der späteren Ergebnisse sicherzustellen. Das vorliegende systematische Review hatte sich zum Ziel gesetzt, die externe Validität von gematchten Propensity-Score-Analysen zu prüfen. Zu diesem Zweck wurden die Ausprägungen der Merkmale Alter, linksventrikuläre Ejektionsfraktion, Geschlecht, Diabetes, Hypertonie und Myokardinfarktgeschichte aus Studien ausgelesen, die ihrerseits Zielgrößen von On- bzw. Off-Pump-Bypass-Operationen untersuchten. Es wurden 38 Propensity-Score-Studien und zum Vergleich ebenfalls 91 randomisiert-kontrollierte Studien eingeschlossen. Für metrische Merkmale wurde der Mittelwert  $\bar{x}_{MA}$  und für binäre Merkmale die Häufigkeit  $\hat{p}_{MA}$  in jeder Studienpopulation ermittelt. Als Vergleichspunkt und anzunehmendes Maß für den Durchschnitt der Bypass-Patienten wurden die Grundgesamtheiten der PS-Studien gemittelt. Es wurden zwei Analysen angefertigt: die sog. Between-Analyse, die die Merkmalsausprägung in Grundgesamtheit, RCT und gematchter Population miteinander verglich, und die sog. Within-Analyse, bei der betrachtet wurde, wie sich die Merkmalsausprägung nach Matches in denjenigen PS-Studien veränderte, die sowohl Angaben zur Grundgesamtheit als auch zu gematchter Population machten. Um zu überprüfen, welchen Einfluss das gewählte statistische Modell und verschiedene Grundannahmen über die Beschaffenheit der Daten auf die berechneten Mittelwerte und Häufigkeiten hat, wurde eine Sensitivitätsanalyse für die Between-Analyse mit einer einfacheren statistischen Modellierung unter Vernachlässigung einiger wichtiger Annahmen durchgeführt.

Bezüglich der externen Validität von RCTs stellte sich heraus, dass die Teilnehmer im Vergleich zur Grundgesamtheit eine signifikant und deutlich erhöhte Ejektionsfraktion hatten ( $\Delta_{General-RCT} = -6,9$ ; [95%KI : -10,5; -3,2]), die die externe Validität beeinträchtigt. Der Anteil der Diabetiker und Hypertoniker unterschied sich kaum. Die Patienten waren in dieser Analyse 1,5 Jahre jünger ( $\Delta_{General-RCT} = 1,5$ ; [95%KI : -0,5; 3,4]), 3,5% mehr Patienten hatten bereits einen Myokardinfarkt ( $\Delta_{General-RCT} = -3,5$ ; [95%KI : -14,5; 7,5]) und es waren 1,4% mehr männliche Probanden ( $\Delta_{General-RCT} = -1,4$ ; [95%KI : -6,6; 3,8]) vorhanden als in der Grundgesamtheit. Abgesehen von der Ejektionsfraktion sind die gefundenen Differenzen jedoch nicht klinisch relevant. Bei der Analyse der Merkmalsausprägungen in gematchten PS-Studien im Vergleich zur Grundgesamtheit fielen keine großen, aber mehrere kleine Differenzen auf. So waren die Teilnehmer im Schnitt 1,3 Jahre älter ( $\Delta_{General-RCT} = -1,3$ ; [95%KI : -1,8; -0,8]), hatten eine um 1,2% erhöhte linksventrikuläre Ejektionsfraktion ( $\Delta_{General-RCT} = -1,2$ ; [95%KI : -2,3; -0,2]), und wiesen einen 2,6 Prozentpunkte niedrigeren Anteil an Männern auf ( $\Delta_{General-RCT} = 2,6$ ; [95%KI : 2,2; 3,1]). Bezüglich des Anteils der Diabetiker und Patienten mit Myokardinfarktgeschichte besteht kaum ein Unterschied, die Differenz von 1,4% mehr Hypertonikern ( $\Delta_{General-RCT} = -1,4$ ;

[95%KI : -3,1;0,4]) ist nicht signifikant. Die externe Validität wird durch die gefundenen Differenzen nicht negativ beeinflusst. In der Sensitivitätsanalyse konnte bestätigt werden, dass die gefundenen Merkmalsdifferenzen auf die Eigenschaften der Studienpopulationen zurückzuführen sind und nicht von der zu Grunde gelegten statistischen Modellierung abhängig sind.

In der Within-Analyse wurden die Merkmalsausprägungen vor und nach Matchen miteinander verglichen. Es bestätigten sich die in der Between-Analyse gefundenen Differenzen. Insgesamt scheinen besonders junge Männer mit niedrigem EF schwer zu matchen zu sein und werden so nicht in das Studienkollektiv aufgenommen. Die externe Validität wird nicht in klinisch relevantem Maße eingeschränkt.

Zusammenfassend ist die externe Validität von gematchten Propensity-Score-Analysen nicht eingeschränkt, da durch Matchen die durchschnittlichen Eigenschaften der ursprünglichen Grundgesamtheit nicht in klinisch relevantem Maße verändert werden. Die randomisiert-kontrollierten Studien sind den gematchten Propensity-Score-Studien bezüglich der externen Validität unterlegen.

## 6 Literaturverzeichnis

- [1] MCKEE, M. ; BRITTON, A. ; BLACK, N. ; MCPHERSON, K. ; SANDERSON, C. ; BAIN, C.: Methods in health services research. Interpreting the evidence: choosing between randomised and non-randomised studies. In: *BMJ* 319 (1999), Nr. 7205, S. 312–5
- [2] ROTHWELL, P. M.: External validity of randomised controlled trials: "to whom do the results of this trial apply?". In: *Lancet* 365 (2005), Nr. 9453, S. 82–93
- [3] HORDIJK-TRION, M. ; LENZEN, M. ; WIJNS, W. ; DE JAEGERE, P. ; SIMOONS, M. L. ; SCHOLTE OP REIMER, W. J. M. ; BERTRAND, M. E. ; MERCADO, N. ; BOERSMA, E. ; EHS-CR INVESTIGATORS: Patients enrolled in coronary intervention trials are not representative of patients in clinical practice: results from the Euro Heart Survey on Coronary Revascularization. In: *Eur Heart J* 27 (2006), Nr. 6, S. 671–8
- [4] COCA, S. G. ; KRUMHOLZ, H. M. ; GARG, A. X. ; PARIKH, C. R.: Underrepresentation of renal disease in randomized controlled trials of cardiovascular disease. In: *JAMA* 296 (2006), Nr. 11, S. 1377–84
- [5] MASOUDI, F. A. ; HAVRANEK, E. P. ; WOLFE, P. ; GROSS, C. P. ; RATHORE, S. S. ; STEINER, J. F. ; ORDIN, D. L. ; KRUMHOLZ, H. M.: Most hospitalized older persons do not meet the enrollment criteria for clinical trials in heart failure. In: *Am Heart J* 146 (2003), Nr. 2, S. 250–7
- [6] VAN SPALL, H. G. C. ; TOREN, A. ; KISS, A. ; FOWLER, R. A.: Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. In: *JAMA* 297 (2007), Nr. 11, S. 1233–40
- [7] SORG, C. ; SCHMIDT, J. ; BÜCHLER, M. W. ; EDLER, L. ; MÄRTEN, A.: Examination of external validity in randomized controlled trials for adjuvant treatment of pancreatic adenocarcinoma. In: *Pancreas* 38 (2009), Nr. 5, S. 542–50
- [8] ADAMINA, M. ; GULLER, U. ; WEBER, W. P. ; OERTLI, D.: Propensity scores and the surgeon. In: *Br J Surg* 93 (2006), Nr. 4, S. 389–94
- [9] ROSENBAUM, P. R. ; RUBIN, D. B.: The central role of the propensity score in observational studies for causal effects. In: *Biometrika* 70 (1983), Nr. 1, S. 41–55
- [10] WINKELMAYER, W. C. ; KURTH, T.: Propensity scores: help or hype? In: *Nephrol Dial Transplant* 19 (2004), Nr. 7, S. 1671–3
- [11] AUSTIN, P. C.: Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement. In: *J Thorac Cardiovasc Surg* 134 (2007), Nr. 5, S. 1128–35
- [12] ROBERT-KOCH-INSTITUT (Hrsg.): *Daten und Fakten: Ergebnisse der Studien "Gesundheit in Deutschland aktuell 2010". Beiträge zur Gesundheitberichterstattung des Bundes.* 2012
- [13] HEROLD, G. (Hrsg.): *Innere Medizin - Eine vorlesungsorientierte Darstellung.* 2013

- [14] STATISTISCHES BUNDESAMT (Hrsg.): *Gesundheit - Fallpauschalenbezogene Krankenhausstatistik (DRG-Statistik) - Operationen und Prozeduren der vollstationären Patientinnen und Patienten in Krankenhäusern*. Wiesbaden, 2011
- [15] HENNE-BRUNS, D. ; KREMER, B. ; DÜRIG, M.: *Duale Reihe - Chirurgie*. Stuttgart : Georg Thieme Verlag KG, 2007
- [16] HENNE-BRUNS, D. ; KREMER, B. ; DÜRIG, M.: *Duale Reihe - Chirurgie*. Stuttgart : Georg Thieme Verlag KG, 2003
- [17] AUSTIN, P. C.: The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. In: *Med Decis Making* 29 (2009), Nr. 6, S. 661–77
- [18] BLACKSTONE, E. H.: Comparing apples and oranges. In: *J Thorac Cardiovasc Surg* 123 (2002), Nr. 1, S. 8–15
- [19] MOLLER, C. H. ; PENNINGA, L. ; WETTERSLEV, J. ; STEINBRUCHEL, D. A. ; GLUUD, C.: Clinical outcomes in randomized trials of off- vs. on-pump coronary artery bypass surgery: systematic review with meta-analyses and trial sequential analyses. In: *Eur Heart J* 29 (2008), Nr. 21, S. 2601–16
- [20] MOLLER, C. H.: Off-pump versus on-pump coronary artery bypass grafting for ischaemic heart disease (Protocol). In: *The Cochrane Library* (2008), Nr. 3
- [21] KUSS, O. ; VON SALVIATI, B. ; BÖRGERMANN, J.: Off-pump versus on-pump coronary artery bypass grafting: a systematic review and meta-analysis of propensity score analyses. In: *J Thorac Cardiovasc Surg* 140 (2010), Nr. 4, S. 829–35, 835.e1–13
- [22] HOZO, S. P. ; DJULBEGOVIC, B. ; HOZO, I.: Estimating the mean and variance from the median, range, and the size of a sample. In: *BMC Med Res Methodol* 5 (2005), S. 13
- [23] SCHABENBERGER, O.: *Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models*. Paper 196-30. SAS Campus Drive, Cary, NC 27513: SAS Institute Inc., 2005
- [24] KUSS, O. ; WATZKE, S.: Korrekter Umgang mit korrelierten Daten in der Rehabilitationsforschung. In: *Rehabilitation 2005*. Georg Thieme Verlag KG, 2005 ( 44), S. 367–372
- [25] HEDEKER, D.: Generalized Linear Mixed Models. In: EVERITT, B. S. (Hrsg.) ; HOWELL, D. (Hrsg.): *Encyclopedia of Statistics in Behavioral Science*. John Wiley and Sons, Ltd, 2005
- [26] SAS INSTITUTE INC.: The MIXED Procedure. In: SAS INSTITUTE INC. (Hrsg.): *SAS/STAT® 9.2 User's Guide*. SAS Campus Drive, Cary, NC 27513, 2008
- [27] BÖRGERMANN, J. ; HAKIM, K. ; KUSS, O. ; SIMM, A. ; SCHEUBEL, R. J. ; BUERKE, M. ; BORGER, M. A. ; GUMMERT, J. F. ; SILBER, R.-E. ; FRIEDRICH, I.: *Comparison between off-pump, on-pump arrested and on-pump beating heart coronary revascularization: Randomized trial in 604 unselected patients (3CAB study)*. – im Review-Prozess (J Thorac Cardiovasc Surg)

- [28] KUSS, O. ; BÖRGERMANN, J.: Do higher-risk patients benefit from off-pump coronary artery bypass grafting? Evidence from an ecologic analysis of randomized trials. In: *J Thorac Cardiovasc Surg* 142 (2011), Nr. 3, S. e117–22

## 7 Anhang

### 7.1 Eingeschlossene Studien

#### 7.1.1 Randomisiert-kontrollierte Studien

- [RCT1] AL-RUZZEH, S. ; GEORGE, S. ; BUSTAMI, M. ; WRAY, J. ; ILSLEY, C. ; ATHANASIOU, T. ; AMRANI, M.: Effect of off-pump coronary artery bypass surgery on clinical, angiographic, neurocognitive, and quality of life outcomes: randomised controlled trial. In: *BMJ* 332 (2006), Nr. 7554, S. 1365
- [RCT2] AL-RUZZEH, S. ; HOARE, G. ; MARCZIN, N. ; ASIMAKOPOULOS, G. ; GEORGE, S. ; TAYLOR, K. ; AMRANI, M.: Off-pump coronary artery bypass surgery is associated with reduced neutrophil activation as measured by the expression of CD11b: a prospective randomized study. In: *Heart Surg Forum* 6 (2003), Nr. 2, S. 89–93
- [RCT3] ALWAN, K. ; FALCOZ, P. E. ; ALWAN, J. ; MOUAWAD, W. ; OUJAIMI, G. ; CHOCRON, S. ; ETIEVENT, J. P.: Beating versus arrested heart coronary revascularization: evaluation by cardiac troponin I release. In: *Ann Thorac Surg* 77 (2004), Nr. 6, S. 2051–5
- [RCT4] ANDERSON, R. E. ; BRISMAR, K. ; BARR, G. ; IVERT, T.: Effects of cardiopulmonary bypass on glucose homeostasis after coronary artery bypass surgery. In: *Eur J Cardiothorac Surg* 28 (2005), Nr. 3, S. 425–30
- [RCT5] ANGELINI, G. D. ; TAYLOR, F. C. ; REEVES, B. C. ; ASCIONE, R.: Early and midterm outcome after off-pump and on-pump surgery in Beating Heart Against Cardioplegic Arrest Studies (BHACAS 1 and 2): a pooled analysis of two randomised controlled trials. In: *Lancet* 359 (2002), Nr. 9313, S. 1194–9
- [RCT6] ASCIONE, R. ; GHOSH, A. ; REEVES, B. C. ; ARNOLD, J. ; POTTS, M. ; SHAH, A. ; ANGELINI, G. D.: Retinal and cerebral microembolization during coronary artery bypass surgery: a randomized, controlled trial. In: *Circulation* 112 (2005), Nr. 25, S. 3833–8
- [RCT7] ASCIONE, R. ; TALPAHEWA, S. ; RAJAKARUNA, C. ; REEVES, B. C. ; LOVELL, A. T. ; COHEN, A. ; ANGELINI, G. D.: Splanchnic organ injury during coronary surgery with or without cardiopulmonary bypass: a randomized, controlled trial. In: *Ann Thorac Surg* 81 (2006), Nr. 1, S. 97–103
- [RCT8] BAKER, R. A. ; ANDREW, M. J. ; ROSS, I. K. ; KNIGHT, J. L.: The Octopus II stabilizing system: biochemical and neuropsychological outcomes in coronary artery bypass surgery. In: *Heart Surg Forum* 4 Suppl 1 (2001), S. S19–23
- [RCT9] BEDNAR, F. ; OSMANCIK, P. ; VANEK, T. ; MOCIKOVA, H. ; JARES, M. ; STRAKA, Z. ; WIDIMSKY, P.: Platelet activity and aspirin efficacy after off-pump compared with on-pump coronary artery bypass surgery: results from the prospective randomized trial PRAGUE 11-Coronary Artery Bypass and REactivity of Thrombocytes (CABARET). In: *J Thorac Cardiovasc Surg* 136 (2008), Nr. 4, S. 1054–60

- [RCT10] BLACHER, C. ; NEUMANN, J. ; JUNG, L. A. ; LUCCHESI, F. A. ; RIBEIRO, J. P.: Off-pump coronary artery bypass grafting does not reduce lymphocyte activation. In: *Int J Cardiol* 101 (2005), Nr. 3, S. 473–9
- [RCT11] BONACCHI, M. ; PRIFTI, E. ; MAIANI, M. ; BARTOLOZZI, F. ; DI EUSANIO, M. ; LEACCHE, M.: Does off-pump coronary revascularization reduce the release of the cerebral markers, S-100beta and NSE? In: *Heart Lung Circ* 15 (2006), Nr. 5, S. 314–9
- [RCT12] CAPUTO, M. ; YEATMAN, M. ; NARAYAN, P. ; MARCHETTO, G. ; ASCIONE, R. ; REEVES, B. C. ; ANGELINI, G. D.: Effect of off-pump coronary surgery with right ventricular assist device on organ function and inflammatory response: a randomized controlled trial. In: *Ann Thorac Surg* 74 (2002), Nr. 6, S. 2088–95; discussion 2095–6
- [RCT13] CARRIER, M. ; PERRAULT, L. P. ; JEANMART, H. ; MARTINEAU, R. ; CARTIER, R. ; PAGE, P.: Randomized trial comparing off-pump to on-pump coronary artery bypass grafting in high-risk patients. In: *Heart Surg Forum* 6 (2003), Nr. 6, S. E89–92
- [RCT14] CASTELLHEIM, A. ; HOEL, T. N. ; VIDEM, V. ; FOSSE, E. ; PHARO, A. ; SVENNEVIG, J. L. ; FIANE, A. E. ; MOLLNES, T. E.: Biomarker profile in off-pump and on-pump coronary artery bypass grafting surgery in low-risk patients. In: *Ann Thorac Surg* 85 (2008), Nr. 6, S. 1994–2002
- [RCT15] CAVALCA, V. ; SISILLO, E. ; VEGLIA, F. ; TREMOLI, E. ; CIGHETTI, G. ; SALVI, L. ; SOLA, A. ; MUSSONI, L. ; BIGLIOLI, P. ; FOLCO, G. ; SALA, A. ; PAROLARI, A.: Isoprostanes and oxidative stress in off-pump and on-pump coronary bypass surgery. In: *Ann Thorac Surg* 81 (2006), Nr. 2, S. 562–7
- [RCT16] CELIK, J. B. ; GORMUS, N. ; TOPAL, A. ; OKESLI, S. ; SOLAK, H.: Effect of off-pump and on-pump coronary artery bypass grafting on renal function. In: *Ren Fail* 27 (2005), Nr. 2, S. 183–8
- [RCT17] COVINO, E. ; SANTISE, G. ; DI LELLO, F. ; DE AMICIS, V. ; BONIFAZI, R. ; BELLINO, I. ; SPAMPINATO, N.: Surgical myocardial revascularization (CABG) in patients with pulmonary disease: beating heart versus cardiopulmonary bypass. In: *J Cardiovasc Surg (Torino)* 42 (2001), Nr. 1, S. 23–6
- [RCT18] CZERNY, M. ; BAUMER, H. ; KILO, J. ; LASSNIGG, A. ; HAMWI, A. ; VUKOVICH, T. ; WOLNER, E. ; GRIMM, M.: Inflammatory response and myocardial injury following coronary artery bypass grafting with or without cardiopulmonary bypass. In: *Eur J Cardiothorac Surg* 17 (2000), Nr. 6, S. 737–42
- [RCT19] CZERNY, M. ; BAUMER, H. ; KILO, J. ; ZUCKERMANN, A. ; GRUBHOFER, G. ; CHEVTCHIK, O. ; WOLNER, E. ; GRIMM, M.: Complete revascularization in coronary artery bypass grafting with and without cardiopulmonary bypass. In: *Ann Thorac Surg* 71 (2001), Nr. 1, S. 165–9
- [RCT20] DIEGELER, A. ; HIRSCH, R. ; SCHNEIDER, F. ; SCHILLING, L. O. ; FALK, V. ; RAUCH, T. ; MOHR, F. W.: Neuromonitoring and neurocognitive outcome in off-pump versus

- conventional coronary bypass operation. In: *Ann Thorac Surg* 69 (2000), Nr. 4, S. 1162–6
- [RCT21] DORMAN, B. H. ; KRATZ, J. M. ; MULTANI, M. M. ; BARON, R. ; FARRAR, E. ; WALTON, S. ; PAYNE, K. ; IKONOMIOIS, J. ; REEVES, S. ; MUKHERJEE, R. ; SPINALE, F. G.: A prospective, randomized study of endothelin and postoperative recovery in off-pump versus conventional coronary artery bypass surgery. In: *J Cardiothorac Vasc Anesth* 18 (2004), Nr. 1, S. 25–9
- [RCT22] ERNEST, C. S. ; WORCESTER, M. U. ; TATOULIS, J. ; ELLIOTT, P. C. ; MURPHY, B. M. ; HIGGINS, R. O. ; LE GRANDE, M. R. ; GOBLE, A. J.: Neurocognitive outcomes in off-pump versus on-pump bypass surgery: a randomized controlled trial. In: *Ann Thorac Surg* 81 (2006), Nr. 6, S. 2105–14
- [RCT23] FATTOUCH, K. ; GUCCIONE, F. ; DIOGUARDI, P. ; SAMPOGNARO, R. ; CORRADO, E. ; CARUSO, M. ; RUVOLO, G.: Off-pump versus on-pump myocardial revascularization in patients with ST-segment elevation myocardial infarction: a randomized trial. In: *J Thorac Cardiovasc Surg* 137 (2009), Nr. 3, S. 650–6; discussion 656–7
- [RCT24] GASZ, B. ; BENKO, L. ; JANCZO, G. ; LANTOS, J. ; SZANTO, Z. ; ALOTTI, N. ; ROTH, E.: Comparison of inflammatory response following coronary revascularization with or without cardiopulmonary bypass. In: *Exp Clin Cardiol* 9 (2004), Nr. 1, S. 26–30
- [RCT25] GEROLA, L. R. ; BUFFOLO, E. ; JASBIK, W. ; BOTELHO, B. ; BOSCO, J. ; BRASIL, L. A. ; BRANCO, J. N.: Off-pump versus on-pump myocardial revascularization in low-risk patients with one or two vessel disease: perioperative results in a multicenter randomized controlled trial. In: *Ann Thorac Surg* 77 (2004), Nr. 2, S. 569–73
- [RCT26] GONENC, A. ; HACISEVKI, A. ; BAKKALOGLU, B. ; SOYAGIR, A. ; TORUN, M. ; KARAGOZ, H. ; SIMSEK, B.: Oxidative stress is decreased in off-pump versus on-pump coronary artery surgery. In: *J Biochem Mol Biol* 39 (2006), Nr. 4, S. 377–82
- [RCT27] GU, Y. J. ; MARIANI, M. A. ; OEVEREN, W. van ; GRANDJEAN, J. G. ; BOONSTRA, P. W.: Reduction of the inflammatory response in patients undergoing minimally invasive coronary artery bypass grafting. In: *Ann Thorac Surg* 65 (1998), Nr. 2, S. 420–4
- [RCT28] GULER, M. ; KIRALI, K. ; TOKER, M. E. ; BOZBUGA, N. ; OMEROGU, S. N. ; AKINCI, E. ; YAKUT, C.: Different CABG methods in patients with chronic obstructive pulmonary disease. In: *Ann Thorac Surg* 71 (2001), Nr. 1, S. 152–7
- [RCT29] GULIELMOS, V. ; MENSCHIKOWSKI, M. ; DILL, H. ; ELLER, M. ; THIELE, S. ; TUGTEKIN, S. M. ; JAROSS, W. ; SCHUELER, S.: Interleukin-1, interleukin-6 and myocardial enzyme response after coronary artery bypass grafting - a prospective randomized comparison of the conventional and three minimally invasive surgical techniques. In: *Eur J Cardiothorac Surg* 18 (2000), Nr. 5, S. 594–601
- [RCT30] HERNANDEZ, Jr. F. ; BROWN, J. R. ; LIKOSKY, D. S. ; CLOUGH, R. A. ; HESS, A. L. ; ROTH, R. M. ; ROSS, C. S. ; WHITED, C. M. ; O'CONNOR, G. T. ; KLEMPERER, J. D.: Neurocognitive outcomes of off-pump versus on-pump coronary artery bypass:

- a prospective randomized controlled trial. In: *Ann Thorac Surg* 84 (2007), Nr. 6, S. 1897–903
- [RCT31] HUEB, W. ; LOPES, N. H. ; PEREIRA, A. C. ; HUEB, A. C. ; SOARES, P. R. ; FAVARATO, D. ; VIEIRA, R. D. ; LIMA, E. G. ; GARZILLO, C. L. ; PAULITCH FDA, S. ; CESAR, L. A. ; GERSH, B. J. ; RAMIRES, J. A.: Five-year follow-up of a randomized comparison between off-pump and on-pump stable multivessel coronary artery bypass grafting. The MASS III Trial. In: *Circulation* 122 (2010), Nr. 11 Suppl, S. S48–52
- [RCT32] JARES, M. ; VANEK, T. ; BEDNAR, F. ; MALY, M. ; SNIRCOVA, J. ; STRAKA, Z.: Off-pump versus on-pump coronary artery surgery. In: *Int Heart J* 48 (2007), Nr. 1, S. 57–67
- [RCT33] JOHANSSON-SYNNERGRÉN, M. ; NILSSON, F. ; BENGTSSON, A. ; JEPPSSON, A. ; WIKLUND, L.: Off-pump CABG reduces complement activation but does not significantly affect peripheral endothelial function: a prospective randomized study. In: *Scand Cardiovasc J* 38 (2004), Nr. 1, S. 53–8
- [RCT34] KHAN, N. E. ; DE SOUZA, A. ; MISTER, R. ; FLATHER, M. ; CLAGUE, J. ; DAVIES, S. ; COLLINS, P. ; WANG, D. ; SIGWART, U. ; PEPPER, J.: A randomized comparison of off-pump and on-pump multivessel coronary-artery bypass surgery. In: *N Engl J Med* 350 (2004), Nr. 1, S. 21–8
- [RCT35] KOBAYASHI, J. ; TASHIRO, T. ; OCHI, M. ; YAKU, H. ; WATANABE, G. ; SATOH, T. ; TAGUSARI, O. ; NAKAJIMA, H. ; KITAMURA, S.: Early outcome of a randomized comparison of off-pump and on-pump multiple arterial coronary revascularization. In: *Circulation* 112 (2005), Nr. 9 Suppl, S. I338–43
- [RCT36] KOCHAMBA, G. S. ; YUN, K. L. ; PFEFFER, T. A. ; SINTEK, C. F. ; KHONSARI, S.: Pulmonary abnormalities after coronary arterial bypass grafting operation: cardiopulmonary bypass versus mechanical stabilization. In: *Ann Thorac Surg* 69 (2000), Nr. 5, S. 1466–70
- [RCT37] KOLACKOVA, M. ; KUDLOVA, M. ; KUNES, P. ; LONSKY, V. ; MANDAK, J. ; ANDRYS, C. ; JANKOVICOVA, K. ; KREJSEK, J.: Early expression of FcγRI (CD64) on monocytes of cardiac surgical patients and higher density of monocyte anti-inflammatory scavenger CD163 receptor in on-pump patients. In: *Mediators Inflamm* 2008 (2008), S. 235461
- [RCT38] LEE, J. D. ; LEE, S. J. ; TSUSHIMA, W. T. ; YAMAUCHI, H. ; LAU, W. T. ; POPPER, J. ; STEIN, A. ; JOHNSON, D. ; LEE, D. ; PETROVITCH, H. ; DANG, C. R.: Benefits of off-pump bypass on neurologic and clinical morbidity: a prospective randomized trial. In: *Ann Thorac Surg* 76 (2003), Nr. 1, S. 18–25; discussion 25–6
- [RCT39] LEGARE, J. F. ; BUTH, K. J. ; KING, S. ; WOOD, J. ; SULLIVAN, J. A. ; HANCOCK FRIESEN, C. ; LEE, J. ; STEWART, K. ; HIRSCH, G. M.: Coronary bypass surgery performed off pump does not result in lower in-hospital morbidity than coronary artery bypass grafting performed on pump. In: *Circulation* 109 (2004), Nr. 7, S. 887–92

- [RCT40] LI, J.-Y. ; LAI, Y.-J. ; YEH, H.-I. ; CHEN, C.-L. ; SUN, S. ; WU, S.-J. ; LIN, F.-Y.: Atrial gap junctions, NF-kappaB and fibrosis in patients undergoing coronary artery bypass surgery: the relationship with postoperative atrial fibrillation. In: *Cardiology* 112 (2009), Nr. 2, S. 81–8
- [RCT41] LINGAAS, P. S. ; HOL, P. K. ; LUNDBLAD, R. ; REIN, K. A. ; TONNESEN, T. I. ; SVENNEVIG, J. L. ; HAUGE, S. N. ; VATNE, K. ; FOSSE, E.: Clinical and Angiographic Outcome of Coronary Surgery with and without Cardiopulmonary Bypass: A Prospective Randomized Trial. In: *Heart Surg Forum* 7 (2004), Nr. 1, S. 37–41
- [RCT42] MALIK, V. ; KALE, S. C. ; CHOWDHURY, U. K. ; RAMAKRISHNAN, L. ; CHAUHAN, S. ; KIRAN, U.: Myocardial injury in coronary artery bypass grafting: On-pump versus off-pump comparison by measuring heart-type fatty-acid-binding protein release. In: *Tex Heart Inst J* 33 (2006), Nr. 3, S. 321–7
- [RCT43] MANDAK, J. ; POJAR, M. ; CIBICEK, N. ; LONSKY, V. ; PALICKA, V. ; KAKRDOVA, D. ; NEDVIDKOVA, J. ; KUBICEK, J. ; ZIVNY, P.: Impact of cardiopulmonary bypass on peripheral tissue metabolism and microvascular blood flow. In: *Perfusion* 23 (2008), Nr. 6, S. 339–46
- [RCT44] MANTOVANI, V. ; KENNERGREN, C. ; BUGGE, M. ; SALA, A. ; LONNROTH, P. ; BERGLIN, E.: Myocardial metabolism assessed by microdialysis: a prospective randomized study in on- and off-pump coronary bypass surgery. In: *Int J Cardiol* 143 (2010), Nr. 3, S. 302–8
- [RCT45] MARISCALCO, G. ; ENGSTROM, K. G. ; FERRARESE, S. ; COZZI, G. ; BRUNO, V. D. ; SESSA, F. ; SALA, A.: Relationship between atrial histopathology and atrial fibrillation after coronary bypass surgery. In: *J Thorac Cardiovasc Surg* 131 (2006), Nr. 6, S. 1364–72
- [RCT46] MASOUMI, M. ; SAIDI, M. R. ; ROSTAMI, F. ; SEPAHI, H. ; ROUSHANI, D.: Off-pump coronary artery bypass grafting in left ventricular dysfunction. In: *Asian Cardiovasc Thorac Ann* 16 (2008), Nr. 1, S. 16–20
- [RCT47] MATATA, B. M. ; SOSNOWSKI, A. W. ; GALINANES, M.: Off-pump bypass graft operation significantly reduces oxidative stress and inflammation. In: *Ann Thorac Surg* 69 (2000), Nr. 3, S. 785–91
- [RCT48] MEDVED, I. ; ANIC, D. ; ZRNIC, B. ; OSTRIC, M. ; SAFTIC, I.: Off-pump versus on-pump–intermittent aortic cross clamping–myocardial revascularisation: single center experience. In: *Coll Antropol* 32 (2008), Nr. 2, S. 381–4
- [RCT49] MICHAUX, I. ; FILIPOVIC, M. ; SKARVAN, K. ; SCHNEITER, S. ; SCHUMANN, R. ; ZERKOWSKI, H. R. ; BERNET, F. ; SEEBERGER, M. D.: Effects of on-pump versus off-pump coronary artery bypass graft surgery on right ventricular function. In: *J Thorac Cardiovasc Surg* 131 (2006), Nr. 6, S. 1281–8
- [RCT50] MODINE, T. ; ZANNIS, C. ; SALLERON, J. ; PROVOT, F. ; GOURLAY, T. ; DUHAMEL, A. ; KOUSSA, M. ; FAYAD, G.: A prospective randomized study to evaluate the

- renal impact of surgical revascularization strategy in diabetic patients. In: *Interact Cardiovasc Thorac Surg* 11 (2010), Nr. 4, S. 406–10
- [RCT51] MOLLER, C. H. ; PERKO, M. J. ; LUND, J. T. ; ANDERSEN, L. W. ; KELBAEK, H. ; MADSEN, J. K. ; WINKEL, P. ; GLUUD, C. ; STEINBRUCHEL, D. A.: No major differences in 30-day outcomes in high-risk patients randomized to off-pump versus on-pump coronary bypass surgery: the best bypass surgery trial. In: *Circulation* 121 (2010), Nr. 4, S. 498–504
- [RCT52] MOTALLEBZADEH, R. ; BLAND, J. M. ; MARKUS, H. S. ; KASKI, J. C. ; JAHANGIRI, M.: Health-related quality of life outcome after on-pump versus off-pump coronary artery bypass graft surgery: a prospective randomized study. In: *Ann Thorac Surg* 82 (2006), Nr. 2, S. 615–9
- [RCT53] MOTALLEBZADEH, R. ; KANAGASABAY, R. ; BLAND, M. ; KASKI, J. C. ; JAHANGIRI, M.: S100 protein and its relation to cerebral microemboli in on-pump and off-pump coronary artery bypass surgery. In: *Eur J Cardiothorac Surg* 25 (2004), Nr. 3, S. 409–14
- [RCT54] MUNERETTO, C. ; BISLERI, G. ; NEGRI, A. ; MANFREDI, J. ; METRA, M. ; NODARI, S. ; DEI CAS, L.: Off-pump coronary artery bypass surgery technique for total arterial myocardial revascularization: a prospective randomized study. In: *Ann Thorac Surg* 76 (2003), Nr. 3, S. 778–82; discussion 783
- [RCT55] NESHER, N. ; FROLKIS, I. ; VARDI, M. ; SHEINBERG, N. ; BAKIR, I. ; CASELMAN, F. ; PEVNI, D. ; BEN-GAL, Y. ; SHARONY, R. ; BOLOTIN, G. ; LOBERMAN, D. ; URETZKY, G. ; WEINBROUM, A. A.: Higher levels of serum cytokines and myocardial tissue markers during on-pump versus off-pump coronary artery bypass surgery. In: *J Card Surg* 21 (2006), Nr. 4, S. 395–402
- [RCT56] NIRANJAN, G. ; ASIMAKOPOULOS, G. ; KARAGOUNIS, A. ; COCKERILL, G. ; THOMPSON, M. ; CHANDRASEKARAN, V.: Effects of cell saver autologous blood transfusion on blood loss and homologous blood transfusion requirements in patients undergoing cardiac surgery on- versus off-cardiopulmonary bypass: a randomised trial. In: *Eur J Cardiothorac Surg* 30 (2006), Nr. 2, S. 271–7
- [RCT57] NOGUEIRA, C. R. ; HUEB, W. ; TAKIUTI, M. E. ; GIRARDI, P. B. ; NAKANO, T. ; FERNANDES, F. ; PAULITSCH FDA, S. ; GOIS, A. F. ; LOPES, N. H. ; STOLF, N. A.: Quality of life after on-pump and off-pump coronary artery bypass grafting surgery. In: *Arq Bras Cardiol* 91 (2008), Nr. 4, S. 217–22, 238–44
- [RCT58] NOUR EL-DIN, B. M.: Clinical Evaluation of Oxygen Transport in Patients Undergoing Off-Pump Compared to Conventional Coronary Artery Bypass Grafting. In: *Eg J Anaesth* (2004), Nr. 20, S. 351–355
- [RCT59] ONORATI, F. ; RUBINO, A. S. ; NUCERA, S. ; FOTI, D. ; SICA, V. ; SANTINI, F. ; GULLETTA, E. ; RENZULLI, A.: Off-pump coronary artery bypass surgery versus standard linear or pulsatile cardiopulmonary bypass: endothelial activation and inflammatory response. In: *Eur J Cardiothorac Surg* 37 (2010), Nr. 4, S. 897–904

- [RCT60] ORHAN, G. ; SARGIN, M. ; SENAY, S. ; YUKSEL, M. ; KURC, E. ; TASDEMIR, M. ; OZAY, B. ; AKA, S. A.: Systemic and myocardial inflammation in traditional and off-pump cardiac surgery. In: *Tex Heart Inst J* 34 (2007), Nr. 2, S. 160–5
- [RCT61] OZKARA, C. ; GULER, N. ; BATYRALIEV, T. ; OKUT, H. ; AGIRBASLI, M.: Does off-pump coronary artery bypass surgery reduce secretion of plasminogen activator inhibitor-1? In: *Int J Clin Pract* 61 (2007), Nr. 5, S. 763–7
- [RCT62] PAROLARI, A. ; ALAMANNI, F. ; JULIANO, G. ; POLVANI, G. ; ROBERTO, M. ; VEGLIA, F. ; FUMERO, A. ; CARLUCCI, C. ; RONA, P. ; BRAMBILLASCA, C. ; SISILLO, E. ; BIGLIOLI, P.: Oxygen metabolism during and after cardiac surgery: role of CPB. In: *Ann Thorac Surg* 76 (2003), Nr. 3, S. 737–43; discussion 743
- [RCT63] PAROLARI, A. ; CAMERA, M. ; ALAMANNI, F. ; NALIATO, M. ; POLVANI, G. L. ; AGRIFOGLIO, M. ; BRAMBILLA, M. ; BIANCARDI, C. ; MUSSONI, L. ; BIGLIOLI, P. ; TREMOLI, E.: Systemic inflammation after on-pump and off-pump coronary bypass surgery: a one-month follow-up. In: *Ann Thorac Surg* 84 (2007), Nr. 3, S. 823–8
- [RCT64] PAULITSCH, F. S. ; SCHNEIDER, D. ; SOBEL, B. E. ; RACHED, R. ; RAMIRES, J. ; JATENE, F. ; STOLF, N. ; HUEB, W. ; LOPES, N. H.: Hemostatic changes and clinical sequelae after on-pump compared with off-pump coronary artery bypass surgery: a prospective randomized study. In: *Coron Artery Dis* 20 (2009), Nr. 2, S. 100–5
- [RCT65] PENTTILA, H. J. ; LEPOJARVI, M. V. ; KIVILUOMA, K. T. ; KAUKORANTA, P. K. ; HASSINEN, I. E. ; PEUHKURINEN, K. J.: Myocardial preservation during coronary surgery with and without cardiopulmonary bypass. In: *Ann Thorac Surg* 71 (2001), Nr. 2, S. 565–71
- [RCT66] POJAR, M. ; MAND'AK, J. ; CIBICEK, N. ; LONSKY, V. ; DOMINIK, J. ; PALICKA, V. ; KUBICEK, J.: Peripheral tissue metabolism during off-pump versus on-pump coronary artery bypass graft surgery: the microdialysis study. In: *Eur J Cardiothorac Surg* 33 (2008), Nr. 5, S. 899–905
- [RCT67] PUSKAS, J. D. ; WILLIAMS, W. H. ; DUKE, P. G. ; STAPLES, J. R. ; GLAS, K. E. ; MARSHALL, J. J. ; LEIMBACH, M. ; HUBER, P. ; GARAS, S. ; SAMMONS, B. H. ; MCCALL, S. A. ; PETERSEN, R. J. ; BAILEY, D. E. ; CHU, H. ; MAHONEY, E. M. ; WEINTRAUB, W. S. ; GUYTON, R. A.: Off-pump coronary artery bypass grafting provides complete revascularization with reduced myocardial injury, transfusion requirements, and length of stay: a prospective randomized comparison of two hundred unselected patients undergoing off-pump versus conventional coronary artery bypass grafting. In: *J Thorac Cardiovasc Surg* 125 (2003), Nr. 4, S. 797–808
- [RCT68] QUANIERS, J. M. ; LERUTH, J. ; ALBERT, A. ; LIMET, R. R. ; DEFRAIGNE, J. O.: Comparison of inflammatory responses after off-pump and on-pump coronary surgery using surface modifying additives circuit. In: *Ann Thorac Surg* 81 (2006), Nr. 5, S. 1683–90
- [RCT69] RACHWALIK, M. ; LYSENKO, L. ; KUSTRYCKI, W. ; PELCZAR, M. ; WACHNIK, J.: Comparison of spirometry in patients after on-pump and off-pump coronary artery bypass grafting. In: *Anaesth Int Therapy* (2006), Nr. 38, S. 4–7

- [RCT70] RAINIO, A. ; HAUTALA, N. ; PELKONEN, O. ; PALOSAARI, T. ; HEIKKINEN, J. ; MOSORIN, M. ; LAHTINEN, J. ; TASKINEN, P. ; ANTTILA, V. ; SURCEL, H. M. ; LEPOJARVI, M. ; JUVONEN, T. ; BIANCARI, F.: Risk of retinal microembolism after off-pump and on-pump coronary artery bypass surgery. In: *J Cardiovasc Surg (Torino)* 48 (2007), Nr. 6, S. 773–9
- [RCT71] RAJA, S. G. ; HAIDER, Z. ; AHMAD, M.: Predictors of gastrointestinal complications after conventional and beating heart coronary surgery. In: *Surgeon* 1 (2003), Nr. 4, S. 221–8
- [RCT72] RASMUSSEN, B. S. ; LAUGESEN, H. ; SOLLID, J. ; GRONLUND, J. ; REES, S. E. ; TOFT, E. ; GJEDSTED, J. ; DETHLEFSEN, C. ; TONNESEN, E.: Oxygenation and release of inflammatory mediators after off-pump compared with after on-pump coronary artery bypass surgery. In: *Acta Anaesthesiol Scand* 51 (2007), Nr. 9, S. 1202–10
- [RCT73] SAHLMAN, A. ; AHONEN, J. ; NEMLANDER, A. ; SALMENPERA, M. ; ERIKSSON, H. ; RAMO, J. ; VENTO, A.: Myocardial metabolism on off-pump surgery; a randomized study of 50 cases. In: *Scand Cardiovasc J* 37 (2003), Nr. 4, S. 211–5
- [RCT74] SAJJA, L. R. ; MANNAM, G. ; CHAKRAVARTHI, R. M. ; SOMPALLI, S. ; NAIDU, S. K. ; SOMARAJU, B. ; PENUMATSA, R. R.: Coronary artery bypass grafting with or without cardiopulmonary bypass in patients with preoperative non-dialysis dependent renal insufficiency: a randomized study. In: *J Thorac Cardiovasc Surg* 133 (2007), Nr. 2, S. 378–88
- [RCT75] SCHMID, F. X. ; VUDATTU, N. ; FLOERCHINGER, B. ; HILKER, M. ; EISSNER, G. ; HOENICKA, M. ; HOLLER, E. ; BIRNBAUM, D. E.: Endothelial apoptosis and circulating endothelial cells after bypass grafting with and without cardiopulmonary bypass. In: *Eur J Cardiothorac Surg* 29 (2006), Nr. 4, S. 496–500
- [RCT76] SCHULZE, C. ; CONRAD, N. ; SCHUTZ, A. ; EGI, K. ; REICHENSPURNER, H. ; REICHART, B. ; WILDHIRT, S. M.: Reduced expression of systemic proinflammatory cytokines after off-pump versus conventional coronary artery bypass grafting. In: *Thorac Cardiovasc Surg* 48 (2000), Nr. 6, S. 364–9
- [RCT77] SELVANAYAGAM, J. B. ; PETERSEN, S. E. ; FRANCIS, J. M. ; ROBSON, M. D. ; KARDOS, A. ; NEUBAUER, S. ; TAGGART, D. P.: Effects of off-pump versus on-pump coronary surgery on reversible and irreversible myocardial injury: a randomized trial using cardiovascular magnetic resonance imaging and biochemical markers. In: *Circulation* 109 (2004), Nr. 3, S. 345–50
- [RCT78] SHROYER, A. L. ; GROVER, F. L. ; HATTLER, B. ; COLLINS, J. F. ; McDONALD, G. O. ; KOZORA, E. ; LUCKE, J. C. ; BALTZ, J. H. ; NOVITZKY, D.: On-pump versus off-pump coronary-artery bypass surgery. In: *N Engl J Med* 361 (2009), Nr. 19, S. 1827–37
- [RCT79] SNIRCOVA, J. ; JARES, M. ; MALY, M. ; STRAKA, Z. ; SPEGAR, J. ; VANEK, T.: Postoperative blood loss in coronary surgery. No real impact of fibrinolysis detected by thromboelastography and D-dimers. A prospective, randomized study. In: *Int Heart J* 49 (2008), Nr. 1, S. 25–38

- [RCT80] SOUSA UVA, M. ; CAVACO, S. ; OLIVEIRA, A. G. ; MATIAS, F. ; SILVA, C. ; MESQUITA, A. ; AGUIAR, P. ; BAU, J. ; PEDRO, A. ; MAGALHAES, M. P.: Early graft patency after off-pump and on-pump coronary bypass surgery: a prospective randomized study. In: *Eur Heart J* 31 (2010), Nr. 20, S. 2492–9
- [RCT81] STRAKA, Z. ; WIDIMSKY, P. ; JIRASEK, K. ; STROS, P. ; VOTAVA, J. ; VANEK, T. ; BRUCEK, P. ; KOLESAR, M. ; SPACEK, R.: Off-pump versus on-pump coronary surgery: final results from a prospective randomized study PRAGUE-4. In: *Ann Thorac Surg* 77 (2004), Nr. 3, S. 789–93
- [RCT82] TANG, A. T. ; KNOTT, J. ; NANSON, J. ; HSU, J. ; HAW, M. P. ; OHRI, S. K.: A prospective randomized study to evaluate the renoprotective action of beating heart coronary surgery in low risk patients. In: *Eur J Cardiothorac Surg* 22 (2002), Nr. 1, S. 118–23
- [RCT83] TATOULIS, J. ; RICE, S. ; DAVIS, P. ; GOLDBLATT, J. C. ; MARASCO, S.: Patterns of postoperative systemic vascular resistance in a randomized trial of conventional on-pump versus off-pump coronary artery bypass graft surgery. In: *Ann Thorac Surg* 82 (2006), Nr. 4, S. 1436–44
- [RCT84] VAN DIJK, D. ; NIERICH, A. P. ; JANSEN, E. W. ; NATHOE, H. M. ; SUYKER, W. J. ; DIEPHUIS, J. C. ; BOVEN, W. J. ; BORST, C. ; BUSKENS, E. ; GROBBEE, D. E. ; ROBLES DE MEDINA, E. O. ; JAEGERE, P. P.: Early outcome after off-pump versus on-pump coronary bypass surgery: results from a randomized study. In: *Circulation* 104 (2001), Nr. 15, S. 1761–6
- [RCT85] VEDIN, J. ; NYMAN, H. ; ERICSSON, A. ; HYLANDER, S. ; VAAGE, J.: Cognitive function after on or off pump coronary artery bypass grafting. In: *Eur J Cardiothorac Surg* 30 (2006), Nr. 2, S. 305–10
- [RCT86] VELISSARIS, T. ; TANG, A. ; MURRAY, M. ; EL-MINSHAWY, A. ; HETT, D. ; OHRI, S.: A prospective randomized study to evaluate splanchnic hypoxia during beating-heart and conventional coronary revascularization. In: *Eur J Cardiothorac Surg* 23 (2003), Nr. 6, S. 917–24; discussion 924
- [RCT87] VELISSARIS, T. ; TANG, A. T. ; WOOD, P. J. ; HETT, D. A. ; OHRI, S. K.: Thyroid function during coronary surgery with and without cardiopulmonary bypass. In: *Eur J Cardiothorac Surg* 36 (2009), Nr. 1, S. 148–54
- [RCT88] VURAL, K. M. ; TASDEMIR, O. ; KARAGOZ, H. ; EMIR, M. ; TARCAN, O. ; BAYAZIT, K.: Comparison of the early results of coronary artery bypass grafting with and without extracorporeal circulation. In: *Thorac Cardiovasc Surg* 43 (1995), Nr. 6, S. 320–5
- [RCT89] WANDSCHNEIDER, W. ; THALMANN, M. ; TRAMPITSCH, E. ; ZIERVOGEL, G. ; KOBINIA, G.: Off-pump coronary bypass operations significantly reduce S100 release: an indicator for less cerebral damage? In: *Ann Thorac Surg* 70 (2000), Nr. 5, S. 1577–9

- [RCT90] ZAMVAR, V. ; WILLIAMS, D. ; HALL, J. ; PAYNE, N. ; CANN, C. ; YOUNG, K. ; KARTHIKEYAN, S. ; DUNNE, J.: Assessment of neurocognitive impairment after off-pump and on-pump techniques for coronary artery bypass graft surgery: prospective randomised controlled trial. In: *BMJ* 325 (2002), Nr. 7375, S. 1268

### 7.1.2 Propensity-Score-Studien

- [PS1] ALAMANNI, F. ; DAINESE, L. ; NALIATO, M. ; GREGU, S. ; AGRIFOGLIO, M. ; POLVANI, G. L. ; BIGLIOLI, P. ; PAROLARI, A.: On- and off-pump coronary surgery and perioperative myocardial infarction: an issue between incomplete and extensive revascularization. In: *Eur J Cardiothorac Surg* 34 (2008), Nr. 1, S. 118–26
- [PS2] ATTARAN, S. ; SHAW, M. ; BOND, L. ; PULLAN, M. D. ; FABRI, B. M.: Does off-pump coronary artery revascularization improve the long-term survival in patients with ventricular dysfunction? In: *Interact Cardiovasc Thorac Surg* 11 (2010), Nr. 4, S. 442–6
- [PS3] BIANCARI, F. ; MAHAR, M. A. ; MOSORIN, M. ; HEIKKINEN, J. ; POKELA, M. ; TASKINEN, P. ; ANTTILA, V. ; LAHTINEN, J. ; LEPOJARVI, M.: Immediate and intermediate outcome after off-pump and on-pump coronary artery bypass surgery in patients with unstable angina pectoris. In: *Ann Thorac Surg* 86 (2008), Nr. 4, S. 1147–52
- [PS4] BOENING, A. ; FRIEDRICH, C. ; HEDDERICH, J. ; SCHOETTLER, J. ; FRAUND, S. ; CREMER, J. T.: Early and medium-term results after on-pump and off-pump coronary artery surgery: a propensity score analysis. In: *Ann Thorac Surg* 76 (2003), Nr. 6, S. 2000–6
- [PS5] BRIZZIO, M. E. ; ZAPOLANSKI, A. ; SHAW, R. E. ; SPERLING, J. S. ; MINDICH, B. P.: Stroke-related mortality in coronary surgery is reduced by the off-pump approach. In: *Ann Thorac Surg* 89 (2010), Nr. 1, S. 19–23
- [PS6] CALAFIORE, A. M. ; DI GIAMMARCO, G. ; TEODORI, G. ; IACO, A. L. ; PANO, M. ; CONTINI, M. ; VITOLLA, G. ; DI MAURO, M.: Bilateral internal thoracic artery grafting with and without cardiopulmonary bypass: six-year clinical outcome. In: *J Thorac Cardiovasc Surg* 130 (2005), Nr. 2, S. 340–5
- [PS7] CALAFIORE, A. M. ; DI MAURO, M. ; CANOSA, C. ; DI GIAMMARCO, G. ; IACO, A. L. ; CONTINI, M.: Early and late outcome of myocardial revascularization with and without cardiopulmonary bypass in high risk patients (EuroSCORE > or = 6). In: *Eur J Cardiothorac Surg* 23 (2003), Nr. 3, S. 360–7
- [PS8] CALAFIORE, A. M. ; DI MAURO, M. ; CANOSA, C. ; DI GIAMMARCO, G. ; IACO, A. L. ; CONTINI, M.: Myocardial revascularization with and without cardiopulmonary bypass: advantages, disadvantages and similarities. In: *Eur J Cardiothorac Surg* 24 (2003), Nr. 6, S. 953–60
- [PS9] CHUKWUEMEKA, A. ; WEISEL, A. ; MAGANTI, M. ; NETTE, A. F. ; WIJEYSUNDERA, D. N. ; BEATTIE, W. S. ; BORGER, M. A.: Renal dysfunction in high-risk patients after

on-pump and off-pump coronary artery bypass surgery: a propensity score analysis. In: *Ann Thorac Surg* 80 (2005), Nr. 6, S. 2148–53

- [PS10] CRESCENZI, G. ; LANDONI, G. ; ROMANO, A. ; BOROLI, F. ; GIARDINA, G. ; BIGNAMI, E. ; FOCHI, O. ; ALETTI, G. ; ROSICA, C. ; ZANGRILLO, A.: A propensity score analysis on the effect of eliminating cardiopulmonary bypass for coronary artery bypass grafting. In: *Minerva Anesthesiol* 73 (2007), Nr. 3, S. 135–41
- [PS11] DI MAURO, M. ; GAGLIARDI, M. ; IACO, A. L. ; CONTINI, M. ; BIVONA, A. ; BOSCO, P. ; GALLINA, S. ; CALAFIORE, A. M.: Does off-pump coronary surgery reduce post-operative acute renal failure? The importance of preoperative renal function. In: *Ann Thorac Surg* 84 (2007), Nr. 5, S. 1496–502
- [PS12] FRANKEL, T. L. ; STAMOU, S. C. ; LOWERY, R. C. ; KAPETANAKIS, E. I. ; HILL, P. C. ; HAILE, E. ; CORSO, P. J.: Risk factors for hemorrhage-related reexploration and blood transfusion after conventional versus coronary revascularization without cardiopulmonary bypass. In: *Eur J Cardiothorac Surg* 27 (2005), Nr. 3, S. 494–500
- [PS13] GORKI, H. ; PATEL, N. C. ; PANAGOPOULOS, G. ; JENNINGS, J. ; BALACUMARASWAMI, L. ; PLESTIS, K. ; SUBRAMANIAN, V. A.: Off-pump Coronary Bypass Surgery in Patients With Low Ejection Fraction: Is There a Long-Term Survival Advantage? In: *Innovations (Phila)* 5 (2010), Nr. 1, S. 33–41
- [PS14] GURU, V. ; GLASGOW, K. W. ; FREMES, S. E. ; AUSTIN, P. C. ; TEOH, K. ; TU, J. V.: The real-world outcomes of off-pump coronary artery bypass surgery in a public health care system. In: *Can J Cardiol* 23 (2007), Nr. 4, S. 281–6
- [PS15] HIX, J. K. ; THAKAR, C. V. ; KATZ, E. M. ; YARED, J. P. ; SABIK, J. ; PAGANINI, E. P.: Effect of off-pump coronary artery bypass graft surgery on postoperative acute kidney injury and mortality. In: *Crit Care Med* 34 (2006), Nr. 12, S. 2979–83
- [PS16] HOLLENBEAK, C. S. ; MORRIS, D. L. ; SINCLAIR, M. C.: Is Off-Pump Coronary Artery Bypass Graft Surgery Cost-Saving? In: *Seminars in Cardiothoracic and Vascular Anesthesia* (2002), Nr. 6, S. 325–329
- [PS17] HU, S. ; ZHENG, Z. ; YUAN, X. ; WANG, W. ; SONG, Y. ; SUN, H. ; XU, J.: Increasing long-term major vascular events and resource consumption in patients receiving off-pump coronary artery bypass: a single-center prospective observational study. In: *Circulation* 121 (2010), Nr. 16, S. 1800–8
- [PS18] KUDUVALLI, M. ; NEWALL, N. ; STOTT, A. ; GRAYSON, A. D. ; FABRI, B. M.: Impact of avoiding cardiopulmonary bypass for coronary surgery on perioperative cardiac enzyme release and survival. In: *Eur J Cardiothorac Surg* 29 (2006), Nr. 5, S. 729–35
- [PS19] LAHTINEN, J. ; BIANCARI, F. ; RIMPILAINEN, J. ; KYTOKORPI, R. ; MOSORIN, M. ; RAINIO, P. ; CRESTI, R. ; JUVONEN, T. ; LEPOJARVI, M.: Off-pump versus on-pump coronary artery bypass surgery in high-risk patients (EuroSCORE  $\geq$  6). In: *Thorac Cardiovasc Surg* 55 (2007), Nr. 1, S. 13–8

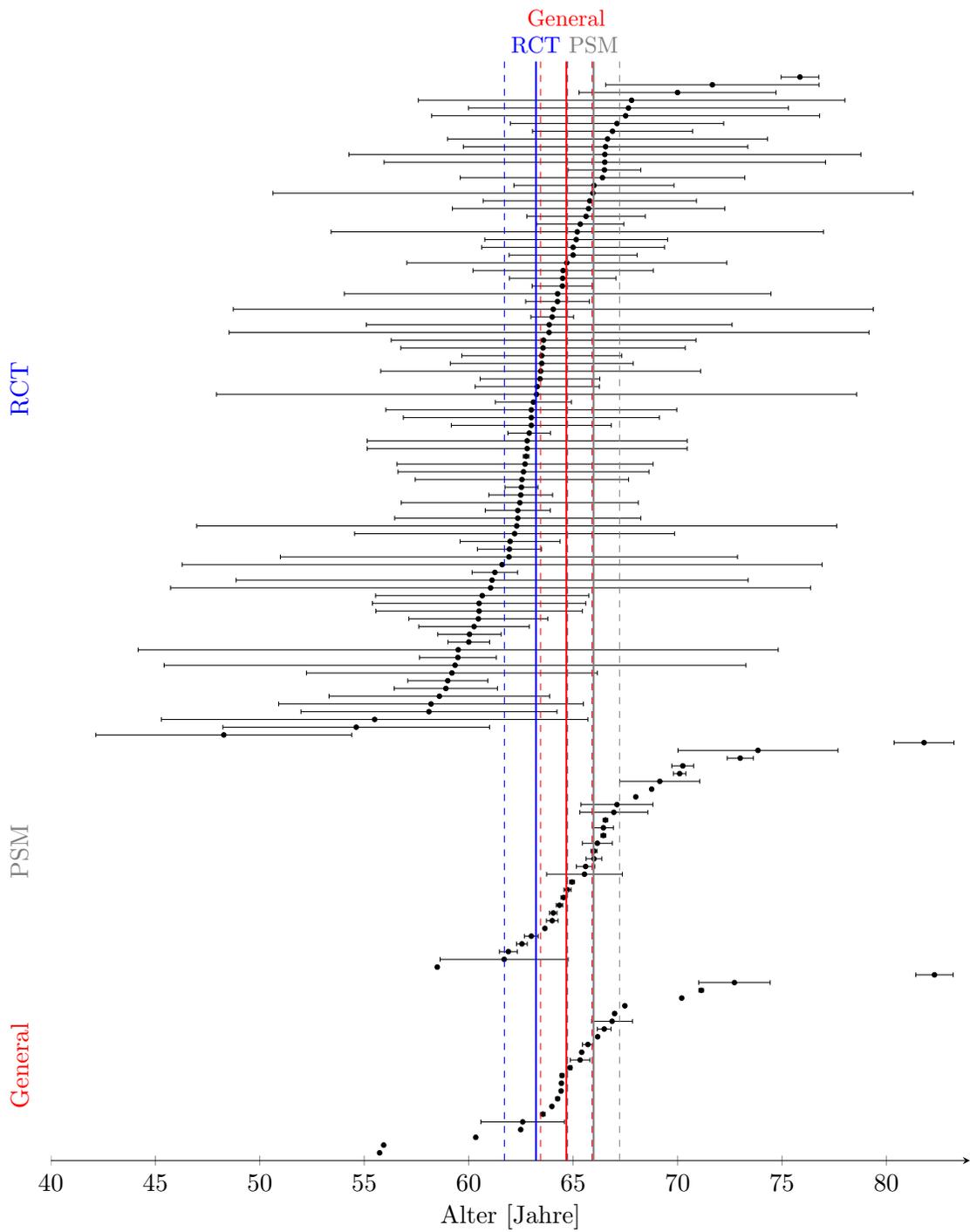
- [PS20] LAMY, A. ; FARROKHYAR, F. ; KENT, R. ; WANG, X. ; SMITH, K. M. ; MULLEN, J. C. ; CARRIER, M. ; CHEUNG, A. ; BAILLOT, R.: The Canadian off-pump coronary artery bypass graft registry: a one-year prospective comparison with on-pump coronary artery bypass grafting. In: *Can J Cardiol* 21 (2005), Nr. 13, S. 1175–81
- [PS21] LOUAGIE, Y. A. ; JAMART, J. ; GRUSLIN, A.: Do coronary bypass graft flows differ between on-pump and off-pump operations? In: *Ann Thorac Surg* 79 (2005), Nr. 6, S. 2004–12
- [PS22] MACK, M. J. ; BROWN, P. ; HOUSER, F. ; KATZ, M. ; KUGELMASS, A. ; SIMON, A. ; BATTAGLIA, S. ; TARKINGTON, L. ; CULLER, S. ; BECKER, E.: On-pump versus off-pump coronary artery bypass surgery in a matched sample of women: a comparison of outcomes. In: *Circulation* 110 (2004), Nr. 11 Suppl 1, S. III–6
- [PS23] MAGANTI, M. ; RAO, V. ; CUSIMANO, R. J.: Sex differences in resource use after on-pump and off-pump coronary artery bypass surgery: a propensity score-matched cohort. In: *Can J Cardiol* 25 (2009), Nr. 5, S. e151–6
- [PS24] MAGEE, M. J. ; COOMBS, L. P. ; PETERSON, E. D. ; MACK, M. J.: Patient selection and current practice strategy for off-pump coronary artery bypass surgery. In: *Circulation* 108 Suppl 1 (2003), S. II9–14
- [PS25] MAGEE, M. J. ; JABLONSKI, K. A. ; STAMOU, S. C. ; PFISTER, A. J. ; DEWEY, T. M. ; DULLUM, M. K. ; EDGERTON, J. R. ; PRINCE, S. L. ; ACUFF, T. E. ; CORSO, P. J. ; MACK, M. J.: Elimination of cardiopulmonary bypass improves early survival for multivessel coronary artery bypass patients. In: *Ann Thorac Surg* 73 (2002), Nr. 4, S. 1196–202
- [PS26] MECO, M. ; BIRAGHI, T. ; PANISI, P. ; CASSELMAN, F. ; COSSETA, D. ; CIRRI, S.: Aortocoronary bypass grafting in high-risk patients over 75 years. Propensity score analysis of on versus off-pump, early and midterm results. In: *J Cardiovasc Surg (Torino)* 48 (2007), Nr. 3, S. 339–47
- [PS27] MISHRA, M. ; MALHOTRA, R. ; KARLEKAR, A. ; MISHRA, Y. ; TREHAN, N.: Propensity case-matched analysis of off-pump versus on-pump coronary artery bypass grafting in patients with atheromatous aorta. In: *Ann Thorac Surg* 82 (2006), Nr. 2, S. 608–14
- [PS28] PANDEY, R. ; GRAYSON, A. D. ; PULLAN, D. M. ; FABRI, B. M. ; DIHMIS, W. C.: Total arterial revascularisation: effect of avoiding cardiopulmonary bypass on in-hospital mortality and morbidity in a propensity-matched cohort. In: *Eur J Cardiothorac Surg* 27 (2005), Nr. 1, S. 94–8
- [PS29] REDDY, S. L. ; GRAYSON, A. D. ; OO, A. Y. ; PULLAN, M. D. ; POONACHA, T. ; FABRI, B. M.: Does off-pump surgery offer benefit in high respiratory risk patients? A respiratory risk stratified analysis in a propensity-matched cohort. In: *Eur J Cardiothorac Surg* 30 (2006), Nr. 1, S. 126–31
- [PS30] SABIK, J. F. ; GILLINOV, A. M. ; BLACKSTONE, E. H. ; VACHA, C. ; HOUGHTALING, P. L. ; NAVIA, J. ; SMEDIRA, N. G. ; MCCARTHY, P. M. ; COSGROVE, D. M. ; LYTLE,

- B. W.: Does off-pump coronary surgery reduce morbidity and mortality? In: *J Thorac Cardiovasc Surg* 124 (2002), Nr. 4, S. 698–707
- [PS31] SALEH, H. Z. ; SHAW, M. ; FABRI, B. M. ; CHALMERS, J. A.: Does avoidance of cardiopulmonary bypass confer any benefits in octogenarians undergoing coronary surgery? In: *Interact Cardiovasc Thorac Surg* 12 (2011), Nr. 3, S. 435–9
- [PS32] SEIF, J. S. ; ZAKY, S. S. ; DEOGAONKAR, A. ; ABD-ELSAIED, A. A. ; LI, M. P. ; LEVENTHAL, M. ; BASHOUR, C. A.: A propensity-matched comparison of the incidence of new-onset postoperative atrial arrhythmias in cardiac and noncardiac thoracic surgery patients. In: *J Cardiothorac Vasc Anesth* 24 (2010), Nr. 3, S. 447–50
- [PS33] SHARONY, R. ; GROSSI, E. A. ; SAUNDERS, P. C. ; GALLOWAY, A. C. ; APPLEBAUM, R. ; RIBAKOVE, G. H. ; CULLIFORD, A. T. ; KANCHUGER, M. ; KRONZON, I. ; COLVIN, S. B.: Propensity case-matched analysis of off-pump coronary artery bypass grafting in patients with atheromatous aortic disease. In: *J Thorac Cardiovasc Surg* 127 (2004), Nr. 2, S. 406–13
- [PS34] STAMOU, S. C. ; HILL, P. C. ; HAILE, E. ; PRINCE, S. ; MACK, M. J. ; CORSO, P. J.: Clinical outcomes of nonelective coronary revascularization with and without cardiopulmonary bypass. In: *J Thorac Cardiovasc Surg* 131 (2006), Nr. 1, S. 28–33
- [PS35] STAMOU, S. C. ; JABLONSKI, K. A. ; GARCIA, J. M. ; BOYCE, S. W. ; BAFI, A. S. ; CORSO, P. J.: Operative mortality after conventional versus coronary revascularization without cardiopulmonary bypass. In: *Eur J Cardiothorac Surg* 26 (2004), Nr. 3, S. 549–53
- [PS36] STAMOU, S. C. ; JABLONSKI, K. A. ; PFISTER, A. J. ; HILL, P. C. ; DULLUM, M. K. ; BAFI, A. S. ; BOYCE, S. W. ; PETRO, K. R. ; CORSO, P. J.: Stroke after conventional versus minimally invasive coronary artery bypass. In: *Ann Thorac Surg* 74 (2002), Nr. 2, S. 394–9
- [PS37] YOUN, Y. N. ; CHANG, B. C. ; HONG, Y. S. ; KWAK, Y. L. ; YOO, K. J.: Early and mid-term impacts of cardiopulmonary bypass on coronary artery bypass grafting in patients with poor left ventricular dysfunction: a propensity score analysis. In: *Circ J* 71 (2007), Nr. 9, S. 1387–94

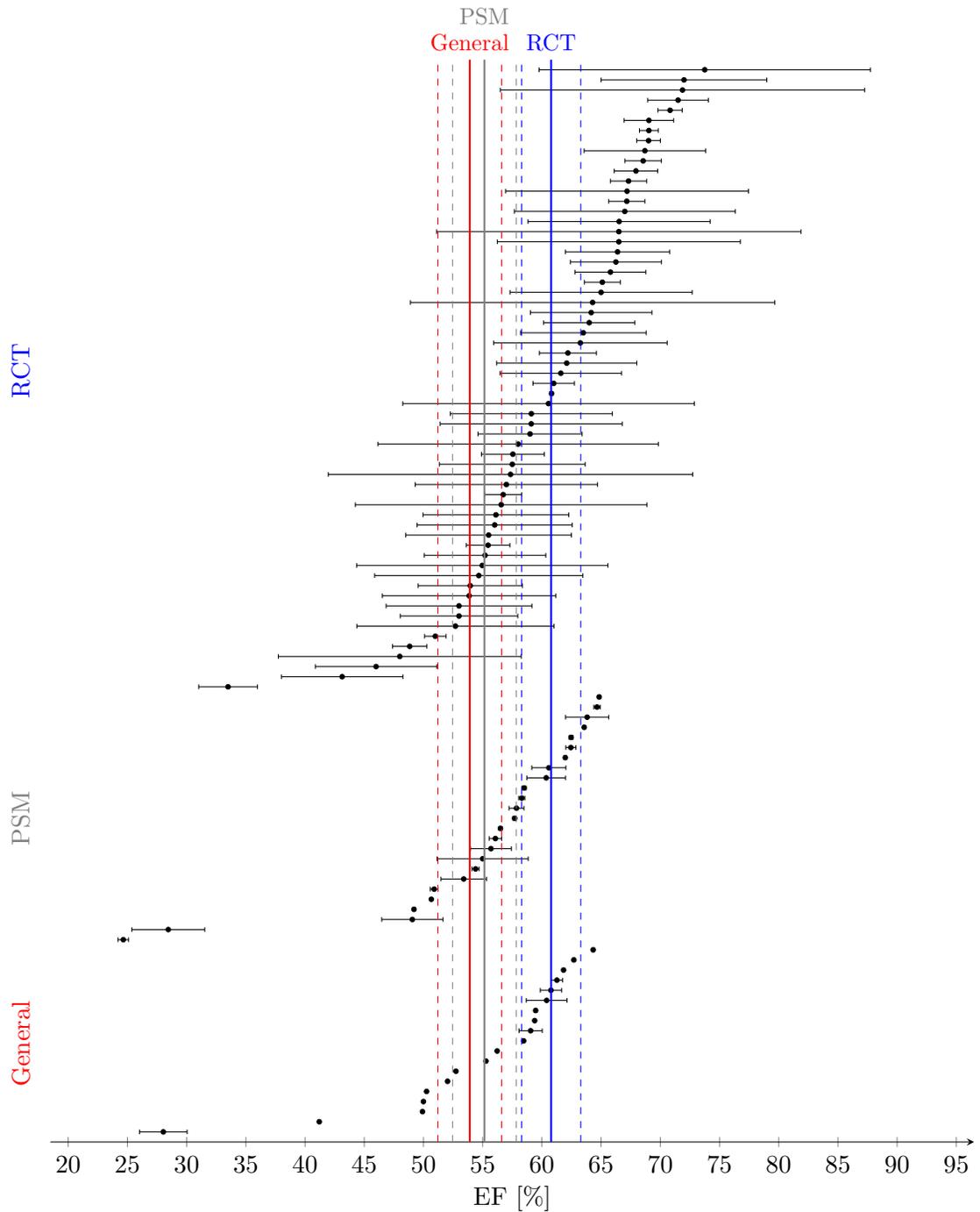
## 7.2 Tabellen und Grafiken

**Tabelle 6:** Dargestellt ist für jede Studienpopulation und für jedes Merkmal, wie viele Daten aus den untersuchten Studien vollständig übernommen werden konnten, wie viele imputiert werden mussten, und in wie vielen Studien Angaben zum jeweiligen Merkmal fehlten.

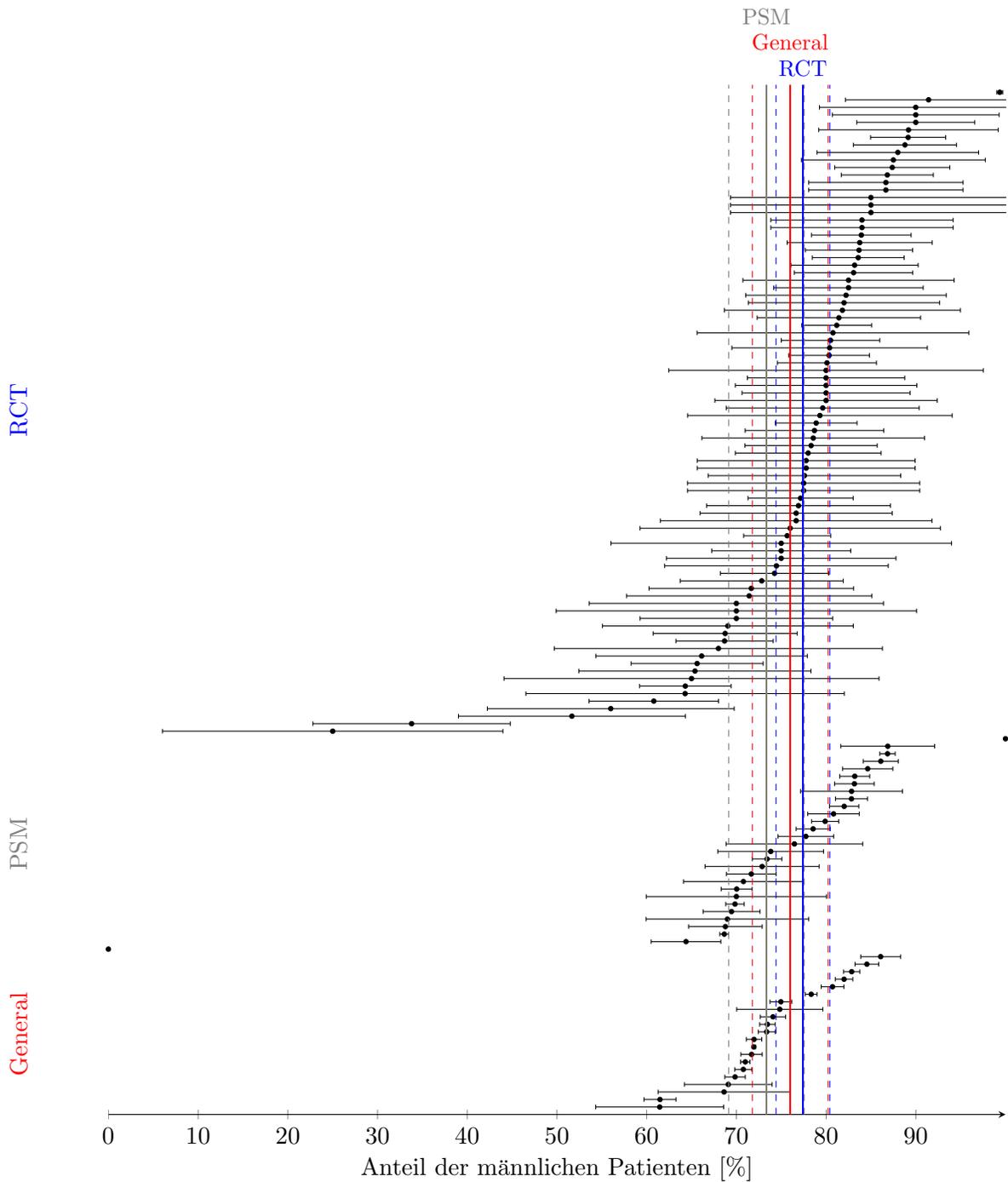
	Variable	gegeben			gesamt
		vollständig	imputiert	fehlend	
RCT	$\bar{x}_{Alter}, s_{Alter}$	78	8	5	91
	$\bar{x}_{EF}, s_{EF}$	35	27	29	91
	$\hat{P}_{Maennlich}$		86	5	91
	$\hat{P}_{Diabetes}$		55	36	91
	$\hat{P}_{Hypertonie}$		58	33	91
	$\hat{P}_{MI}$		30	61	91
	PSM	$\bar{x}_{Alter}, s_{Alter}$	22	8	1
$\bar{x}_{EF}, s_{EF}$		12	13	6	31
$\hat{P}_{Maennlich}$			29	2	31
$\hat{P}_{Diabetes}$			27	4	31
$\hat{P}_{Hypertonie}$			17	14	31
$\hat{P}_{MI}$			18	13	31
General		$\bar{x}_{Alter}, s_{Alter}$	13	11	0
	$\bar{x}_{EF}, s_{EF}$	4	15	5	24
	$\hat{P}_{Maennlich}$		21	3	24
	$\hat{P}_{Diabetes}$		19	5	24
	$\hat{P}_{Hypertonie}$		12	12	24
	$\hat{P}_{MI}$		10	14	24



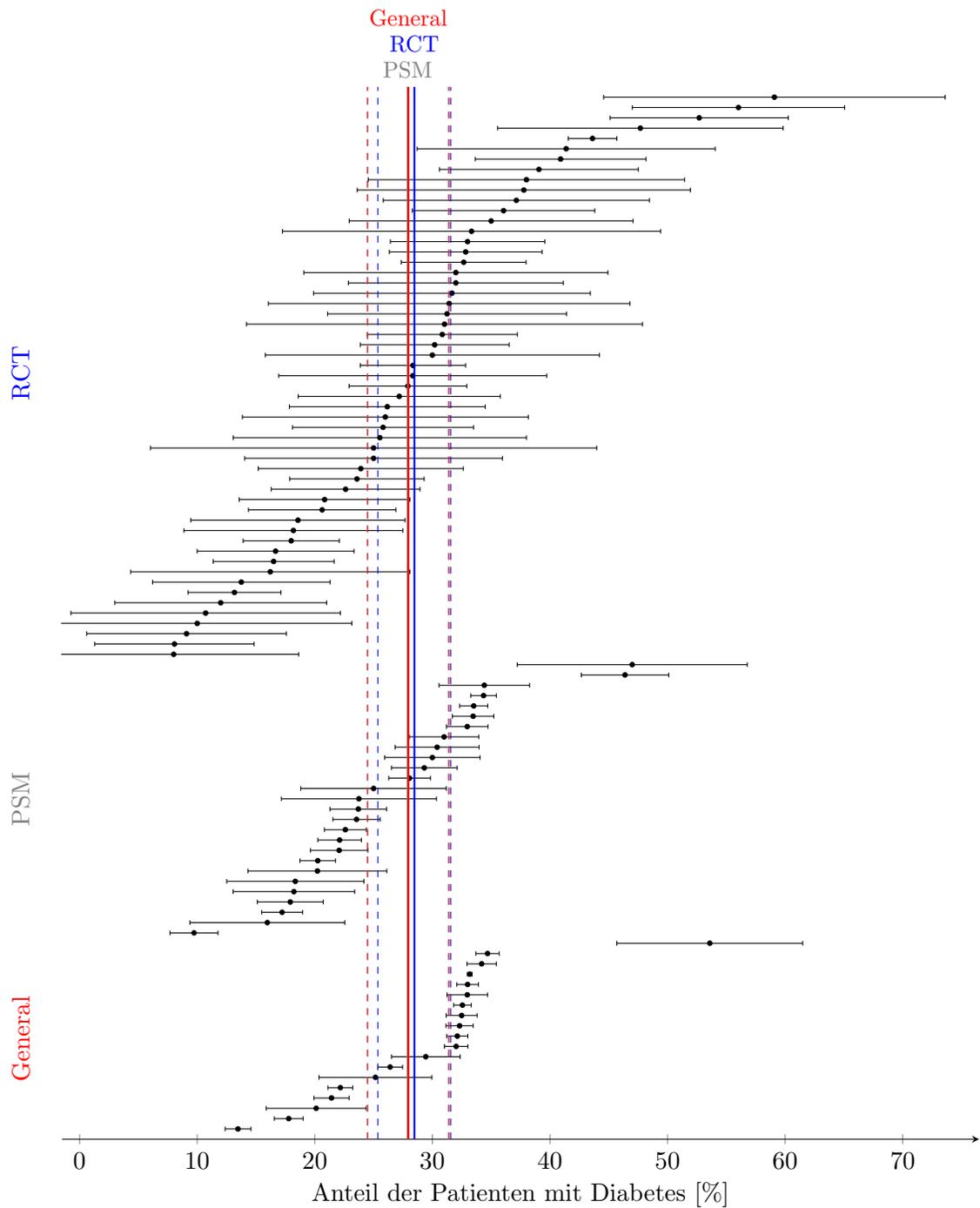
**Abbildung 21:** Mittelwerte  $\bar{x}$  des Alters der Patienten aller analysierten Studien, getrennt nach Studienpopulation, geordnet nach Größe von  $\bar{x}$ . Die Breite der Fehlerbalken ist proportional zum inversen Fallzahlgewicht, mit dem die jeweilige Studie in die Berechnung einging (lange Balken  $\hat{=}$  kleines Gewicht et vice versa). Senkrecht dargestellt sind die Meta-Mittelwerte  $\bar{x}_{MA}$  der drei Studienpopulationen.



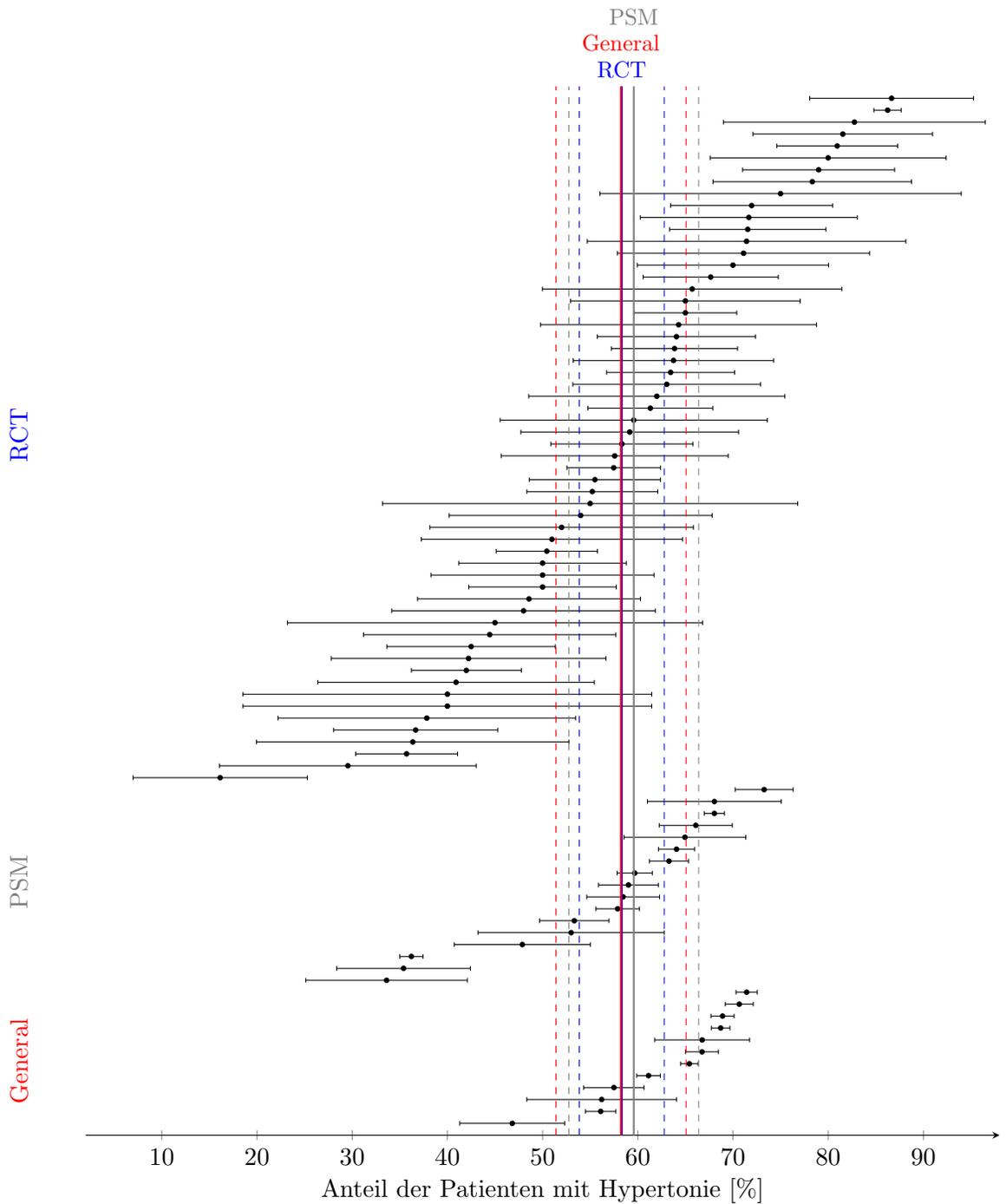
**Abbildung 22:** Mittelwerte  $\bar{x}$  der linksventrikulären Ejektionsfraktion der Patienten aller analysierten Studien, getrennt nach Studienpopulation, geordnet nach Größe von  $\bar{x}$ . Die Breite der Fehlerbalken ist proportional zum inversen Fallzahlgewicht, mit dem die jeweilige Studie in die Berechnung einging (lange Balken  $\hat{=}$  kleines Gewicht et vice versa). Senkrecht dargestellt sind die Meta-Mittelwerte  $\bar{x}_{MA}$  der drei Studienpopulationen.



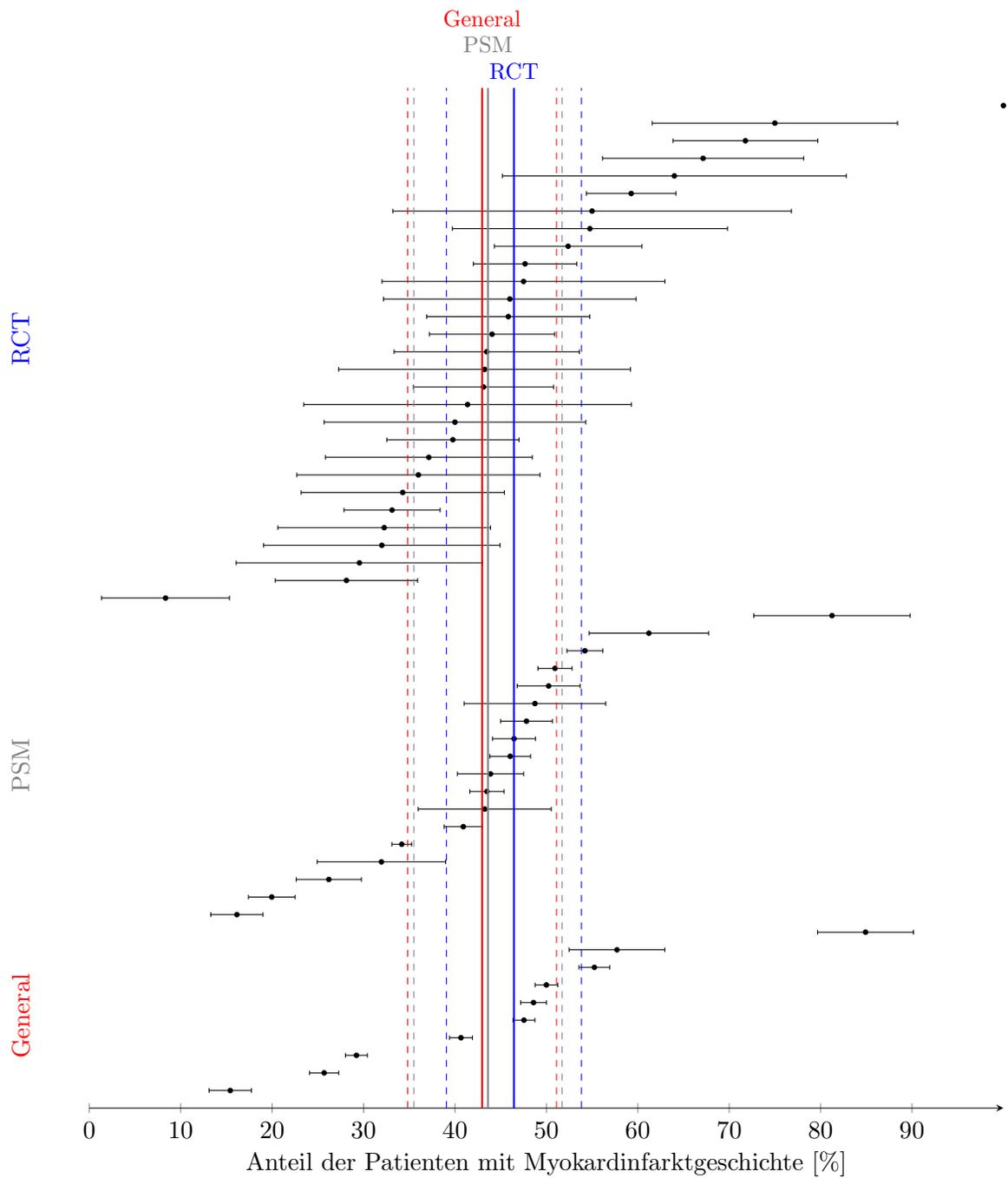
**Abbildung 23:** Häufigkeit  $\hat{p}$  der Probanden männlichen Geschlechts aller analysierten Studien getrennt nach Studienpopulation, geordnet nach Größe der  $\hat{p}$ . Die Fehlerbalken entsprechen hier dem 95%-Konfidenzintervall und somit *nicht* dem exaktem Gewicht der Studien, da in PROC GLIMMIX die Gewichtung implizit vorgenommen wurde. Senkrecht dargestellt sind die Meta-Häufigkeiten  $\hat{p}_{MA}$  der drei Studienpopulationen.



**Abbildung 24:** Häufigkeit  $\hat{p}$  der Diabetiker aller analysierten Studien getrennt nach Studienpopulation, geordnet nach Größe der  $\hat{p}$ . Die Fehlerbalken entsprechen hier dem 95%-Konfidenzintervall und somit *nicht* dem exaktem Gewicht der Studien, da in PROC GLIMMIX die Gewichtung implizit vorgenommen wurde. Senkrecht dargestellt sind die Meta-Häufigkeiten  $\hat{p}_{MA}$  der drei Studienpopulationen.



**Abbildung 25:** Häufigkeit  $\hat{p}$  der Hypertoniker aller analysierten Studien getrennt nach Studienpopulation, geordnet nach Größe der  $\hat{p}$ . Die Fehlerbalken entsprechen hier dem 95%-Konfidenzintervall und somit *nicht* dem exaktem Gewicht der Studien, da in PROC GLIMMIX die Gewichtung implizit vorgenommen wurde. Senkrecht dargestellt sind die Meta-Häufigkeiten  $\hat{p}_{MA}$  der drei Studienpopulationen.



**Abbildung 26:** Häufigkeit  $\hat{p}$  der Probanden mit Myokardinfarktgeschichte aller analysierten Studien getrennt nach Studienpopulation, geordnet nach Größe der  $\hat{p}$ . Die Fehlerbalken entsprechen hier dem 95%-Konfidenzintervall und somit *nicht* dem exaktem Gewicht der Studien, da in PROC GLIMMIX die Gewichtung implizit vorgenommen wurde. Senkrecht dargestellt sind die Meta-Häufigkeiten  $\hat{p}_{MA}$  der drei Studienpopulationen.

## 8 Thesen

1. Zur Prüfung von Therapien sollten in vergleichenden Studien die Behandlungsgruppen so gebildet werden, dass eine möglichst hohe interne und externe Validität der Ergebnisse sichergestellt wird. Eine Studie gilt als intern valide, wenn die gewonnenen Ergebnisse für die Studienpopulation korrekt sind, und als extern valide, wenn die gewonnenen Ergebnisse auf die Gesamtheit aller erkrankten Patienten übertragen werden können.
2. Der derzeitige Goldstandard zur Erstellung von Behandlungsgruppen in Therapiestudien ist die Randomisierung.
3. Randomisiert-kontrollierte Studien verfügen durch Strukturgleichheit über eine hohe interne Validität, während die externe Validität auf Grund rigoroser Auswahl der Patienten oft eingeschränkt ist.
4. Alternativ können Behandlungsgruppen auch mit nicht-randomisierenden Verfahren erstellt werden.
5. 1983 publizierten Rosenbaum und Rubin den Propensity-Score (PS), mit dessen Hilfe man z. B. durch Matchen ebenfalls Behandlungsgruppen erstellen kann. Beim Matchen anhand des PS werden retrospektiv Probanden mit möglichst ähnlichem PS auf die Behandlungsgruppen verteilt. Über viele Probanden gemittelt, ähneln sich Probanden mit ähnlichem Propensity-Score in den zuvor festgesetzten Merkmalen; es wird nur für bekannte Einflussgrößen adjustiert.
6. Gematchte PS-Studien sind bezüglich ihrer externen Validität eingeschränkt, da nicht zu matchende Probanden nicht in das Studienkollektiv aufgenommen werden und bezüglich ihrer internen Validität eingeschränkt, da nur für bekannte Confounder adjustiert werden kann.
7. Bis dato liegen keine wissenschaftlichen Untersuchungen zur externen Validität von gematchten PS-Studien vor.
8. Im vorliegenden systematischen Review wird die externe Validität von 38 gematchten PS-Studien (und zum Vergleich auch die von 91 randomisiert-kontrollierten Studien) geprüft, die ihrerseits verschiedene Zielgrößen bei On- bzw. Off-Pump-Bypass-Operationen untersuchten.
9. Als externer Vergleichspunkt zur Überprüfung der externen Validität, und somit als anzu-nehmender Durchschnitt aller Bypass-Patienten, dienten im vorliegenden systematischen Review die in den Propensity-Score-Studien angegebenen Grundgesamtheiten, also die Population, aus der die Matches gezogen wurden.

10. In der Between-Analyse wurden die durchschnittlichen Eigenschaften der Grundgesamtheit mit denen der gematchten PS-Studienpopulation und RCT-Population verglichen. In der Within-Analyse wurde untersucht, ob und wie durch Matchen die durchschnittlichen Eigenschaften der Grundgesamtheit verändert werden.
11. Es wurde für jede Studienpopulation der Meta-Mittelwert  $\bar{x}_{MA}$  der metrischen Merkmale Alter und linksventrikuläre Ejektionsfraktion bzw. die Meta-Häufigkeit  $\hat{p}_{MA}$  der binären Merkmale Geschlecht, Diabetes, Hypertension und Myokardinfarktgeschichte ermittelt; anschließend wurden die Meta-Lagemaße mit Hilfe des Random-Effects-Modells für metrische Daten und eines Generalized linear mixed models mit Identitätslink für binäre Daten miteinander verglichen.
12. Es lässt sich feststellen, dass die externe Validität von gematchten Propensity-Score-Studien, die Zielgrößen in On- bzw. Off-Pump-Bypass-Operationen miteinander vergleichen, nicht eingeschränkt ist.
13. Die externe Validität von randomisiert-kontrollierten Studien, die Zielgrößen in On- bzw. Off-Pump-Bypass-Operationen miteinander vergleichen, ist bezüglich der linksventrikulären Ejektionsfraktion eingeschränkt; die RCT-Studienpopulation ist hier gesünder als die Grundgesamtheit.
14. Durch Matchen wird ein Kollektiv in Propensity-Score-Studien zu On- bzw. Off-Pump-Bypass-Operationen jünger, hat eine im Gegensatz zur Grundgesamtheit leicht erhöhte linksventrikuläre Ejektionsfraktion und einen höheren Anteil von Frauen. Die externe Validität wird hierdurch nicht in klinisch relevantem Maße eingeschränkt.
15. Die randomisiert-kontrollierten Studien sind den gematchten Propensity-Score-Studien bezüglich ihrer externen Validität unterlegen.
16. Gematchte Propensity-Score-Studien sind in Fachgebieten mit sehr gut erforschten Einflussgrößen eine billigere, weniger aufwendige und extern validere Alternative zu randomisiert-kontrollierten Studien.

## 9 Tabellarischer Lebenslauf

### Schulische Ausbildung

03.07.1989	geboren in Boston (USA)
1995 - 1997	Grundschule Kenzingen
1997 - 2000	Mozart-Grundschule Duisburg
2000 - 2003	Steinbart-Gymnasium Duisburg
2003 - 2005	Carl-Zeiss-Gymnasium Jena
2005 - 2008	Christliches Gymnasium Jena
2008	Abitur (1,0)

### Universitäre Ausbildung

2008 - jetzt	Stipendiatin der Studienstiftung des Deutschen Volkes
2008 - 2010	Humanmedizin-Studium an der Julius-Maximilians-Universität Würzburg
2010	Erster Abschnitt der Ärztlichen Prüfung (2,0)
2010 - jetzt	Humanmedizin-Studium an der Martin-Luther-Universität Halle-Wittenberg

### Praktika und Famulaturen

Aug. 2009	3-wöchiges Forschungspraktikum Biomedizin Universität Lleida, Spanien
Feb. 2010	3-wöchiges Forschungspraktikum Neuroästhetik Friedrich-Schiller-Universität Jena
Jul./Aug. 2010	4-wöchiges Forschungspraktikum Neurophysiologie Julius-Maximilians-Universität Würzburg
Okt./Nov 2011	4-wöchige Famulatur Allgemeinmedizin O <sub>2</sub> -Medical-Network, Cusco, Peru
Nov. 2011 - Feb. 2012	12-wöchige Famulatur Kinderheilkunde, Innere Medizin, Chirurgie Hospital Regional Cusco, Peru

## **10 Selbstständigkeitserklärung**

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte 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 für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Halle, den 11.02.2013

Lisa Angelika Redies

## **11 Erklärung über frühere Promotionsversuche**

Hiermit erkläre ich, dass ich bisher keine früheren Promotionsversuche mit dieser oder einer anderen Dissertation unternommen habe. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Halle, den 11.02.2013

Lisa Angelika Redies

## 12 Danksagung

Herrn PD Dr. Oliver Kuß gilt mein besonderer Dank für die Überlassung des Themas und die Unterstützung bei der Konzeption und Durchführung der Arbeit sowie bei der Erstellung der Dissertationsschrift.

Herrn Prof. Dr. Johannes Haerting und den Mitarbeitern des Instituts für Medizinische Epidemiologie, Biometrie und Informatik möchte ich für wertvolle Denkanstöße anlässlich einer institutsinternen Präsentation des Dissertationsthemas danken.

Auch danke ich meinen Eltern Christoph Redies und Roswitha Koppe-Redies für eine abschließende Rechtschreibkontrolle.