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

Beste verfügbare Evidenz – methodische Herausforderungen und Heterogenitätsanalysen in systematischen Übersichtsarbeiten

Habilitationsschrift

zur Erlangung des akademischen Grades
eines habilitierten Doktors der Medizinischen Wissenschaften (Dr. rer. medic. habil.)
für das Fachgebiet
Medizinische Biometrie und Epidemiologie

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

von Dr.rer.nat. Susanne Unverzagt
geboren am 15.02.1969 in Wernigerode

Gutachter/Gutachterin: Prof. Dr. Antje Timmer, Prof. Dr. Meinhard Kieser
Datum der Vorlesung:12.11.2015
Datum der Verteidigung: 8.12.2015

Referat

Die Validität einer systematischen Übersichtsarbeit wird maßgeblich durch die methodische Qualität der eingeschlossenen Studien und die Konsistenz ihrer Behandlungseffekte bestimmt. Inkonsistente Effekte können neben dem Zufall durch die spezifischen Eigenschaften der Patienten, Interventionen und Endpunkte und die gewählte Studienmethodik bedingt sein. Diese Arbeit beschreibt anhand zweier ausgewählter klinischer Fragestellungen aus der Kardiologie und Allgemeinmedizin die sich daraus ergebenen methodischen Probleme und deren Lösung.

Auf Grundlage des klinisch nicht relevanten Behandlungseffektes einer randomisierten Studie entstand die Idee einer systematischen Zusammenfassung aller Studien zur Wirksamkeit einer intra-aortalen Ballongegenpulsation (IABP) in der Behandlung des infarktbedingten kardiogenen Schocks. Auf ein weiteres Problem in der Versorgung von Patienten mit Herz-Kreislauferkrankungen zielt eine allgemeinmedizinisch motivierte Arbeit zur Wirksamkeit von Implementierungsstrategien auf die Leitlinien- (LL-) Konformität von Ärzten. Methodische Probleme traten in beiden Arbeiten in der Erfassung und Bewertung des Verzerrungspotentials, der Schätzung der Behandlungseffekte und in den anschließenden Heterogenitätsuntersuchungen auf. Die Problembearbeitung erfolgte unter Verwendung hierarchischer Modelle zur Untersuchung des prognostischen Einflusses der Intervention und der Effektmodifikation durch Verzerrungsrisiken und weitere studienspezifische Eigenschaften.

Auf dem Gebiet der Intensiv- und Notfallmedizin konnten wir Evidenz für einen fehlenden Nutzen des Einsatzes einer IABP bei Patienten mit kardiogenen Schock und eine Effektüberschätzung in monozentrischen Studien sammeln. Der Effekt der unterschiedlichen Implementierungsstrategien auf die ärztliche LL-Konformität kann durch klinische Faktoren wie den Adressaten der Strategie, die Patientenpopulation und den Zeitraum, über welchen hinweg diese eingesetzt werden, beeinflusst werden. Aber auch methodische Faktoren wie das Studiendesign und potentielle Verzerrungsquellen können den Behandlungseffekt einer Implementierungsstrategie modifizieren.

Unverzagt, Susanne: Beste verfügbare Evidenz: Methodische Herausforderungen und Heterogenitätsanalysen in systematischen Übersichtsarbeiten, Halle (Saale), Univ. Med. Fak., Habil., 120 Seiten, 2015.

Inhaltsverzeichnis

Verzeichnis der Abkürzungen und Symbole

AHCPR	Agency for Health Care Policy and Research, Department of Health and Human Services
β	Regressionskoeffizient
BMBF	Bundesministerium für Bildung und Forschung
CVD	Herz-Kreislauferkrankungen (engl. „cardiovascular diseases“)
EbM	Evidenzbasierte Medizin
engl.	Englisch
FEM	Fixed-effects-Modell
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HTA	Health Technology Assessment
HR	Hazard Rate (engl. „hazard ratio“)
I	Indikatorvariable
I^2	Quantifiziert die statistische Heterogenität von Behandlungseffekten
IABP	Intraaortalen Ballongegenpulsation (engl. „intra-aortal balloon counter-pulsation”)
ICC	Intracluster-Korrelationskoeffizient (engl. „intracluster correlation coef-ficient”)
IPD	individuelle Patientendaten (engl. „individual patient data”)
KI	Konfidenzintervall
KKSH	Koordinierungszentrum für Klinische Studien Halle
LL	Leitlinien
MAP	Mittlerer arterieller Blutdruck (engl. „mean arterial pressure“)
n	Anzahl
n.b.	Nicht berechnet
OR	Odds Ratio
p	Wahrscheinlichkeit
PROSPERO	Database of Prospectively Registered Systematic Reviews in health and social care
REM	Random-effects-Modell
RCT	randomisierte kontrollierte Studie (engl. „randomized controlled trial“)
ROC	Reciever operating characteristics

ROR	Relative Odds Ratio
SR	Systematischer Review (Übersichtsarbeit)
Tab.	Tabelle
UKH	Universitätsklinikum Halle
vs.	versus
z.B.	zum Beispiel
ZI	Zentralinstitut für die kassenärztliche Versorgung

1. Einleitung

Im Editorial der Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen wurde die Frage „Ist EbM nach 20 Jahren in der Versorgung angekommen?“ sowohl von den Editoren als auch vom Autor mit „Angekommen ja, aber noch nicht heimisch“ (Ollenschläger 2014) beantwortet. Evidenzbasierte Medizin (EbM) ist „der gewissenhafte, ausdrückliche und vernünftige Gebrauch der gegenwärtig besten externen, wissenschaftlichen Evidenz für Entscheidungen in der medizinischen Versorgung individueller Patienten“ (AZQ 2007, Sackett 1996). Neben den Patientenpräferenzen und den persönlichen klinischen Erfahrungen der Ärzte sollen auch kritisch bewertete Studiendaten in der medizinischen Entscheidungsfindung berücksichtigt werden. Der Einsatz der ebM stellt damit die Forderung auf, die persönlichen klinischen Erfahrungen von Ärzten zu hinterfragen und alle eingesetzten medizinischen Verfahren in hochwertigen Studien hinsichtlich ihrer Wirkung, ihres Nutzens und der möglichen Nebenwirkungen für den Patienten zu überprüfen.

Randomisierte kontrollierte Studien (RCTs) können, soweit sie gut geplant, durchgeführt und berichtet werden, Schätzwerte für den Behandlungserfolg einer Intervention mit dem geringsten Verzerrungspotenzial liefern (Chalmers 2001, Schulz 2010). Der anschließende Wissenstransfer der generierten Erkenntnisse in die klinische Praxis setzt eine hohe interne Validität dieser Studien, eine transparente und umfassende Publikation der Studienergebnisse und eine Wissenssynthese in systematischen Übersichtsarbeiten voraus. Diese Arbeiten entstehen in einem reproduzierbaren Prozess, in welchem das zu einer festen Forschungsfrage vorhandene Wissen mit dem Ziel zusammengeführt wird, alle relevanten Daten von Probanden, die jemals an Studien zur Forschungsfrage teilgenommen haben, zu vereinen (Cochrane Collaboration 2011). So kann die Voraussetzung für einen niedrigschwelligen Zugang zur vorhandenen Evidenz für praktisch tätige Ärzte und Patienten über HTA (Health Technology Assessment)-Berichte, klinische Leitlinien und Patienteninformationen (Antes 2014) geschaffen werden.

Systematische Reviews (SR) werden in verschiedenen Evidenzhierarchiesierungen mit der höchsten Evidenzstufe beurteilt (z.B. OCEBM 2011, SIGN 2008). Aber häufig rechtfertigt die zusammengefasste Evidenz (engl. „evidence“ = Aussage, Zeugnis, Beweis) aus diesen Arbeiten keine starke Leitlinienempfehlung für oder gegen die untersuch-

ten Interventionen. Die GRADE-Gruppe empfiehlt eine Herabstufung der Empfehlungsstärke bei eingeschränkter interner Validität, wenn systematisch verfälschte Aussagen aufgrund von Fehlern und Verzerrungen in der Planung und Durchführung der Studie nicht ausgeschlossen werden können. Weitere Gründe für eine reduzierte Evidenz sind unpräzise oder inkonsistente Schätzer, indirekte Evidenz und Publikationsbias (Guyatt 2011a).

Die erste Einschränkung der Zuverlässigkeit der Evidenz eines SRs und der daraus abgeleiteten Behandlungsempfehlungen kann sich aus systematischen Verzerrungen der Behandlungseffekte durch eine fehlerhaften Planung, Durchführung oder Auswertung der Einzelstudien ergeben (Guyatt 2011b), so dass ein wesentlicher Aspekt von Wissenssynthesen in der Kontrolle von Verzerrungsrisiken besteht. Deshalb stellten wir uns die Frage, welche spezifische Studieneigenschaften und systematischen Verzerrungs- (Bias-) quellen den Behandlungseffekt und die darauf basierenden Empfehlungen beeinflussen können (Unverzagt 2013). Wir untersuchten in einer metaepidemiologischen Studie (Definition laut Goodman 2011) Risikofaktoren für eine Effektverzerrung in Metaanalysen, welche durch Eigenschaften der Studien oder Metaanalysen bedingt sind (Unverzagt 2013).

Eine mangelnde Präzision der beobachteten Behandlungseffekte mit einem breiten 95 % Konfidenzintervall (KI), welches sehr unterschiedliche Empfehlungen rechtfertigen würde (Guyatt 2011c), konnten wir in den meist kleinen, monozentrisch durchgeführten Studien der Intensiv- und Notfallmedizin (Unverzagt 2011) regelmäßig feststellen. Aber auch aus der auf den Nachweis einer Mortalitätssenkung gepowerten, abschließend durchgeführten, multizentrischen Studie konnte keine Empfehlung für die Intervention abgeleitet werden (Unverzagt 2015).

Viele der eingeschlossenen Studien mit kleiner Fallzahl wählten zeitnah und metrisch messbare Endpunktdifferenzen als Hauptzielkriterium. Beispiele dafür sind hämodynamische Parameter in der Intensiv- und Notfallmedizin und Prozessparameter wie die Leitlinien- (LL-) Adhärenz oder Konformität eines Arztes. Alle diese Parameter dienen als Surrogate für eine verbesserte Prognose hinsichtlich patientenrelevanter Endpunkte wie dem Gesamtüberleben oder der Lebensqualität der Patienten. Wir empfehlen deshalb, analog zu Guyatt (2011e), eine Abwertung der vorliegenden Evidenz wegen Indirektheit, solange die Assoziation zwischen Surrogatendpunkt und patientenrele-

vantem Endpunkt für die untersuchte Intervention und Indikation nicht eindeutig geklärt ist. Voraussetzung für die Wirksamkeit aller therapeutischen Verfahren und Empfehlungen auf patientenrelevante Endpunkte ist, dass beispielsweise hämodynamischen Verbesserungen des mittleren arteriellen Blutdrucks oder des Herzindex bei Infarktpatienten eine Verringerung der Sterblichkeit bedingen. Auch evidenzbasierte Therapien in der Prävention von Herz-Kreislauferkrankungen (CVD) erfordern neben der in Unverzagt (2014b) untersuchten Adhärenz des Arztes eine Adhärenz des Patienten, so dass Arztadhärenz kein hinreichendes Kriterium für patientenrelevante Endpunkte wie eine verbesserte Morbidität und Mortalität sein kann. LL-Konformität beschreibt dabei die „Übereinstimmung des Wissens, Denkens und Handelns eines Akteurs mit den in einer Leitlinie gegebenen Empfehlungen“ (Hasenbein 2007). Auch bei guter Adhärenz des Arztes entscheiden die Patienten abhängig von ihrem Bildungsniveau, persönlichen Präferenzen, Kostenfaktoren wie Zuzahlungen oder fehlenden Erstattungen, neurologischen oder psychiatrischen Komorbiditäten und der Komplexität des Therapieregimes (Laufs 2011), inwieweit sie den Behandlungsempfehlungen ihrer Ärzte folgen werden.

Inkonsistente Studienergebnisse mit einer hohen Variabilität der geschätzten Behandlungseffekte führten zu einer Abwertung der Qualität der Evidenz in einem SR zur Implementierung von Leitlinien (Unverzagt 2014b). Diese ist dann zweifelhaft, wenn einige Studien bei Anwendung einer Implementierungsstrategie einen substantiellen Nutzen nahelegten, während andere keine Wirksamkeit zeigen (Guyatt 2011d, Perleth 2012). Die statistische Heterogenität lässt sich aus dem Zusammenwirken von klinischen und methodischen Studieneigenschaften erklären (Glasziou 2002, Perleth 2012, Pigott 2013). Die Untersuchung der klinischen Studieneigenschaften bietet die Möglichkeit, Behandlungseffekte besser zu verstehen und die Umgebung zu beschreiben, in welcher eine Intervention am besten wirkt. Diese Untersuchungen beschäftigen sich mit der „externen Validität“ einer Studie (Windeler 2008), und beschreiben die Übertragbarkeit (engl. „generalizability“) der Studienergebnisse. Aber auch methodische Eigenschaften der Einzelstudien wie verschiedene Studiendesigns und das Risiko systematischer Verzerrungen in der Planung und Durchführung der Studien und im Berichten ihrer Ergebnisse können Behandlungseffekte modifizieren und sollten, soweit möglich, kritisch untersucht werden.

2. Zielstellung

In dieser Arbeit möchte ich klinische Ergebnisse und methodische Überlegungen aus der Durchführung einer randomisierten Studie (Prondzinsky 2010), zweier systematischer Übersichtsarbeiten (Unverzagt 2014b und Unverzagt 2015), einer meta-epidemiologischen Studie (Unverzagt 2013) und ausführlicher Heterogenitätsbetrachtungen (Unverzagt 2014c) zusammenfassend darstellen und diskutieren.

Ziel dieser Arbeit ist die Vorstellung von Methoden, wie unter Nutzung aller verfügbaren Informationen und geeigneter Modelle Behandlungseffekte und ihre Modifikation geschätzt und Einschränkungen in der zu einer konkreten medizinischen Fragestellung vorliegenden Evidenz aufgedeckt werden konnten. Daraus ergaben sich sowohl konkrete klinische Handlungsempfehlungen als auch methodische Untersuchungen zu Auswirkungen einer eingeschränkten internen und externen Validität auf die Effektschätzer in SRs.

Einschränkungen der internen Validität in der Planung, Durchführung und anschließenden Beschreibung der Ergebnisse von Einzelstudien können zu systematischen Verzerrungen der Behandlungseffekte in den resultierenden Metaanalysen führen. Deshalb soll hier ein Vorschlag zur Definition potentieller Verzerrungsquellen und zur Beeinflussung des Behandlungseffektes am Beispiel von Studien aus der Intensiv- und Notfallmedizin vorgestellt werden (Unverzagt 2013).

Ein weiteres Problem kann eine hohe unerklärte Variabilität der geschätzten Behandlungseffekte darstellen, welche sowohl innerhalb von Einzelstudien als auch zwischen diesen auftreten kann. Wenn eine ausgeprägte Heterogenität der Behandlungseffekte unerklärt bleibt, sinkt das Vertrauen in die Übertragbarkeit der Behandlungseffekte und damit in die Qualität der Evidenz, so dass auf die Zusammenfassung der Effektschätzer in Metaanalysen verzichtet werden muss. Deshalb werde ich verschiedene Methoden und Ergebnisse aus der statistischen Beurteilung des Ausmaßes von Heterogenität und zur Beschreibung der Effektmodifikation durch klinische und methodische Studieneigenschaften beschreiben (Unverzagt 2014c).

3. Material und Methodik

3.1 Randomisierte klinische Studien am Beispiel der IABP-Shock Studie

Viele methodische Herausforderungen in systematischen Übersichtsarbeiten entstehen aufgrund von Einschränkungen in der methodischen Qualität, der Durchführung und Publikation der Einzelstudien, denn die Qualität eines SRs kann nicht besser als diejenige der Einzelstudien sein (engl. „Garbage in – garbage out.“).

Als Fallbeispiel dafür dient eine monozentrisch durchgeführte RCT (IABP-Shock-Studie) am Universitätsklinikum Halle/Wittenberg (UKH), in welcher ich als Projektleiterin am Koordinierungszentrum für Klinische Studien (KKSH) für die methodische Planung, Durchführung und Veröffentlichung verantwortlich war (Prondzinsky 2010).

Ziel der IABP-SHOCK Studie war die Einschätzung der Wirksamkeit und Sicherheit einer intraaortalen Ballongegenpulsation (IABP) bei Patienten mit infarktbedingtem kardiogenen Schock. Alle Patienten mit infarktbedingten kardiogenem Schock wurden, wenn sie den Ein- und Ausschlusskriterien genügten, unmittelbar nach der notfallmäßigen Einweisung in das UKH in die Studie eingeschlossen und anschließend in einen der beiden Therapiearme randomisiert. Die Randomisierung erfolgte im Herzkatheterlabor auf Grundlage einer blockweisen Randomisierung über das Ziehen durchnummerierter, verschlossener, nicht einsehbarer Briefumschläge, welche im KKSH erstellt wurden. Dabei wurden die Patienten entweder einer Gruppe mit Standardbehandlung ohne Unterstützung durch eine IABP (oder andere mechanische Kreislaufunterstützungsverfahren) oder einer Gruppe mit zusätzlicher IABP-Unterstützung zugewiesen. Die Standardbehandlung umfasste etablierte klinikinterne Behandlungsalgorithmen. Eine Verblindung der Studienteilnehmer und des betreuenden medizinischen Personals wurde aus praktischen Gründen nicht durchgeführt, denn ein IABP-Zugang ist auf einer Intensivstation für Ärzte und die betreuenden Schwestern offensichtlich.

Die für den Nachweis einer verringerten Sterblichkeit notwendige Fallzahl konnten wir aufgrund der innerhalb von nur zwei Jahren am UKH maximal rekrutierbaren 40 Patienten nicht erreichen. Die Durchführung einer multizentrischen Studie hingegen war wegen der hohen Akzeptanz der IABP sowie der sehr hohen Empfehlungsgrade in den nationalen und internationalen Leitlinien zu diesem Zeitpunkt nicht möglich. Deshalb wählten wir die Senkung der Morbidität als Hauptzielkriterium dieser RCT. Zur Quanti-

fizierung der Senkung der Morbidität diente der Apache II-Score während der ersten vier Tage im Krankenhaus. Dieser erlaubt innerhalb von 24 Stunden nach einem akuten Ereignis eine Vorhersage des weiteren Krankheitsverlaufes (Werdan 2007) und basiert auf zwölf Parametern, welche während der klinischen Routineuntersuchungen gemessen werden. Geringere Scorewerte stehen für eine geringere Krankheitslast und bessere Prognose des Patienten, Scorewerte und klinische Prognose sind somit invers korreliert.

Die Nebenzielkriterien dieser Studie umfassten Parameter zur Messung von Veränderungen hämodynamischer, pro- und anti-inflammatorischer Parameter und wurden ebenfalls über vier Tage nach der Randomisierung erfasst.

Alle Schritte zur Datenspeicherung, -kontrolle und -übertragung entsprachen der Planung in einem Handbuch und erfolgten im KKSH. Die ausgefüllten Fragebögen wurden hinsichtlich der Ein- und Ausschlusskriterien, Behandlungszuordnungen, Exposition zur IABP, den Haupt- und Nebenzielzielkriterien, dem Studienende und den Nebenwirkungen durch eine Studienschwester auf der Grundlage der Quelldaten überprüft. Zwei Mitarbeiter des KKSH übertrugen die Daten anschließend unabhängig voneinander in ein Datenbanksystem (Pharma Open Source Community PhOSCo), so dass Eingabe- und Übertragungsfehler verhindert wurden. Nach abgeschlossener Dateneingabe, Datenabgleich und notwendigen Rückfragen und Korrekturen erfolgte der Datenexport der gesamten Datenbank in ein kompatibles Datenformat zur Auswertung in SAS und SPSS für Windows, Version 16 (SPSS Inc., Chicago, IL, USA).

Eine Differenz des Apache-II-Scores von vier Punkten zwischen den Behandlungsgruppen gilt als relevant für die klinische Prognose des Patienten. Eine Feststellung dieser Differenz bei einer Standardabweichung von vier Punkten war bei einer Gruppengröße von zweimal 20 Patienten, der Verwendung des zweiseitigen *t*-Tests und einem Typ I-Fehlers von 5 % mit einer Power von 86 % möglich. Fehlende Werte am Tag 4 nach der Randomisierung traten auf, wenn Patienten frühzeitig verstarben oder sich schnell erholten und zu diesem Zeitpunkt bereits von der Intensivstation entlassen wurden. Diese Werte ersetzten wir konservativ mittels der „last observation carried forward"-Methode.

Alle Analysen basieren auf der „Full analysis" Population der eingeschlossenen Patienten. Patienten wurden aus dieser Population ausgeschlossen, wenn sie die Einschluss-

kriterien nicht erfüllten, aber diese Bewertung zum Zeitpunkt des Einschlusses der Patienten objektiv möglich gewesen wäre oder wenn die Patienten keinerlei studienspezifische Maßnahmen erhielten (ICH E9 1998). Diese Prüfung der Ausschlussgründe erfolgte verblindet gegenüber der Therapiezuweisung.

Neben der Hauptanalyse mit dem zweiseitigen t-Test rechneten wir Sensitivitätsanalysen, um potentielle Verzerrungsquellen und eine daraus resultierende Effektmodifikation zu quantifizieren. Diese beinhalteten eine per-Protokoll-Analyse und Varianzanalysen mit Adjustierungen für zur Randomisierung ungleich verteilter, aber prognostisch wichtiger Variablen.

Zusätzlich untersuchten wir den Einfluss der IABP und die prognostische Bedeutung des mittleren Verlaufes der Haupt- und Nebenzielparameter über die ersten vier Tage nach der Randomisierung und beschrieben diese Veränderungen über die Mittelwerte mit den zugehörigen 95 % KI (Prondzinsky 2010, Prondzinsky 2012a, 2012b, 2012c). Für die pro- und anti-inflammatorischen Zytokine untersuchten wir die prognostische Aussagekraft der initialen und extremsten, während der ersten vier Tage beobachteten, Werte in ROC- und multiplen logistischen Regressionsanalysen (Prondzinsky 2012b, 2012c). Alle Analysen zu den sekundären Endpunkten führten wir hypothesengenerierend durch und verzichteten deshalb auf eine Adjustierung für multiple Tests.

3.2 Systematische Übersichtsarbeiten

Die Durchführung der hier vorgestellten Übersichtsarbeiten (Unverzagt 2014b, Unverzagt 2015) entspricht den AMSTAR-Qualitätskriterien (Shea 2008) und die Veröffentlichungen enthalten alle in den PRISMA-Kriterien (Moher 2009) geforderten Informationen. Die Übersichtsarbeiten sollen hier in der Reihenfolge ihrer Bearbeitung (und Veröffentlichung) kurz vorgestellt werden und können im Volltext in den Anlagen 1 und 3 dieser Arbeit nachgelesen werden.

3.2.1 Intra-aortale Ballongegenpulsation bei Patienten mit infarktbedingtem kardiogenen Schock

Dieser SR (Unverzagt 2011 und im Update 2015) schließt alle RCTs zum Einsatz der IABP als Ergänzung zur Standardtherapie bei Patienten mit infarktbedingtem kardiogenem Schock ein. Wir verglichen Studienarme mit und ohne IABP und ergänzten folgen-

de Vergleiche in Subgruppen in Abhängigkeit von der Behandlung in den Vergleichsgruppen ohne IABP:

1.) Standardtherapie mit IABP vs. dieselbe Standardtherapie ohne IABP oder andere linksventrikuläre Unterstützungssysteme
2.) Standardtherapie mit IABP vs. dieselbe Standardtherapie mit einem anderen linksventrikulären Unterstützungssystem (Impella oder TandemHeart)

Hauptzielkriterien zur Beurteilung der Wirksamkeit (engl. „efficacy“) unter den Idealbedingungen einer RCT sind neben dem Überleben der Patienten als hierarchisch höher gewerteter Endpunkt das Auftreten schwerwiegender nicht-letaler Ereignisse. Das Überleben wurde in den Einzelstudien zu unterschiedlichen Zeitpunkten berichtet, so dass wir im SR Analysen zum Kurz- und Langzeit-Überleben durchführten. Das Kurzzeitüberleben umfasst die Zeit im Krankenhaus oder über 30 Tage, während das Langzeitüberleben die Periode über sechs Monate bis zu einem Jahr nach der Randomisierung umfasst. Die nicht-letalen Ereignisse enthalten erneute Infarkte, Schlaganfälle, rekurrente Ischämien, Wiederverschlüsse der Koronargefäße und die Notwendigkeit einer wiederholten Revaskularisation. Die Nebenzielkriterien umfassen sowohl hämodynamische Parameter als auch die Länge des Aufenthaltes im Krankenhaus und auf der Intensivstation zur Bewertung der Wirksamkeit als auch das Auftreten gerätespezifischer Komplikationen zur Bewertung der Sicherheit der IABP.

Alle Ein- und Ausschlusskriterien, das methodische Vorgehen und die Suchstrategie planten wir in einem in der Cochrane Library publizierten Protokoll. Die Durchführung dieser Arbeit wurde durch das Bundesministerium für Bildung und Forschung (BMBF 01KG0811) gefördert.

3.2.2 Implementierungsstrategien zur leitliniengerechten allgemeinmedizinischen Versorgung von Patienten mit Herz-Kreislauferkrankungen

Dieser SR (Unverzagt 2014b) basiert auf allen individuell- und clusterrandomisiert durchgeführten Studien zum Einsatz von Implementierungsstrategien für LL-Empfehlungen zur allgemeinmedizinischen Behandlung von Patienten mit CVD. Wir verglichen für jede untersuchte Kategorie den Einsatz einer Implementierungsstrategie mit der passiven LL-Implementierung („usual care“):

1.) Unimodale Strategiekategorie vs. Standardversorgung und
2.) Multimodale Implementierungskategorie vs. Standardversorgung.

Alle verwendeten Implementierungsstrategien teilten wir nach Definitionen der Agency for Healthcare Research and Quality (Shojania 2004) in die acht Kategorien Anwender-Erinnerungssysteme, Unterstützung von Datenflüssen, Audit und Feedback, Anwender-/Fortbildungsmaßnahmen (-schulungen), Patientenschulung, Unterstützung des Patienten-Selbstmanagements, Patienten-Erinnerungshilfen und organisatorische Veränderungen im Versorgungsablauf. Das Hauptzielkriterium dieser Arbeit ist die Beurteilung der Wirksamkeit der einzelnen Strategiekategorien auf die LL-Konformität des primärversorgenden Arztes über eine Mindestnachbeobachtungszeit von drei Monaten. Die LL-Konformität des Arztes maßen wir am Anteil LL-konform versorgter Patienten in relevanten Handlungsfeldern wie Beratungen, Diagnostik- und Therapieempfehlungen.

Die Nebenzielkriterien beschreiben die Ergebnisse der Implementierungsstrategien am Patienten und umfassen das Gesamtüberleben, Morbidität und Mortalität durch CVD, individuelle Risikofaktoren, Lebensqualität und Kosten.

Alle Ein- und Ausschlusskriterien und das methodische Vorgehen planten wir in einem in der Online-Datenbank PROSPERO (Database of Prospectively Registered Systematic Reviews in health and social care, Reg. Nr. CRD42011001793) publizierten Protokoll. Die Durchführung dieser Arbeit wurde durch das Zentralinstitut für die kassenärztliche Versorgung in Deutschland (ZI) unterstützt.

3.2.3 Schätzung des Verzerrungsrisiko

Aus dem Cochrane Handbuch (Cochrane Collaboration 2011, Kapitel 8.4.a) übernahmen und bewerteten wir folgende Ursachen für systematisch verfälschte Aussagen (Bias) zur Bewertung der internen Validität der eingeschlossenen RCTs:

- Selektionsbias (engl. „selection bias“), bedingt durch Fehler in der zufälligen Generierung der Zufallsfolge und der verdeckten Therapiezuweisung, beurteilten wir mit „gering“, wenn ein Zufallsprozess zur Generierung der Therapiezuordnung wie computergenerierte Zufallszahlen, Zufallszahlentabellen, Minimierung, Münzwurf oder Würfeln beschrieben wurde und die verdeckte Therapiezuweisung ohne Kenntnis der Randomisierungsgruppe erfolgte. Geeignete Zuweisungsmethoden beinhalten eine zentrale Zuweisung per Telefon, Fax oder verschlossene, blickdichte Briefumschläge.

- Durchführungsbias (engl. „performance bias“), bedingt durch eine unvollständige Verblindung der Studienteilnehmer oder des medizinischen Personals, beurteilten wir mit „gering“, wenn eine doppelte Verblindung gewährleistet war und während der Studiendurchführung aufrechterhalten werden konnte.
- Messungsbias (engl. „detection bias“), bedingt durch Unterschiede in der Messung der Zielkriterien in den Behandlungsgruppen, beurteilten wir mit „gering“, wenn die Messung der Zielkriterien ohne Kenntnis der Randomisierungsgruppe erfolgte oder eine Beeinflussung des Messergebnisses aufgrund von objektiven Kriterien weitgehend ausgeschlossen werden konnte.
- Verlustbias (engl. „attrition bias“), bedingt durch Unterschiede zwischen den Behandlungsgruppen, die in der Nachbeobachtungszeit entstehen (z.B. durch Studienabbrüche und den Ausschluss von Patienten), beurteilten wir mit „gering“, wenn alle randomisierten Patienten in die Analysen eingeschlossen wurden oder die Anzahl der ausgeschiedenen Patienten gering war, gruppenweise Gründe mit Häufigkeiten angegeben wurden und diese in den Behandlungsgruppen vergleichbar waren.
- Publikationsbias (engl. „reporting bias“), bedingt durch Abweichungen zwischen geplanten und berichteten Haupt- und Nebenzielkriterien in Abhängigkeit von den beobachteten Behandlungseffekten, beurteilten wir mit „gering“, wenn alle im Studienprotokoll, in der Studienregistrierung oder im Methodenteil aufgezählten Endpunkte berichtet wurden und der in die Fallzahlanalyse eingeflossene Endpunkt als Hauptzielkriterium berichtet wurde.

Wir nutzten diese Kriterien in den Übersichtsarbeiten, um das Verzerrungsrisiko auf den geschätzten Behandlungseffekt der eingeschlossenen Studien einzuschätzen. Das Verzerrungsrisiko während der Generierung der Zufallsfolge, der verdeckten Therapiezuweisung und des selektiven Berichtens von Endpunkten beurteilten wir spezifisch für jede Studie, während das unvollständige Berichten und die Verblindung für jeden untersuchten Endpunkt bewertet wurde.

In die Übersichtsarbeit zur Untersuchung von Implementierungsstrategien (Unverzagt 2014b) schlossen wir clusterrandomisierte Studien (c-RCT) ein und entwickelten dafür auf der Basis methodischer Empfehlungen von Puffer (2003), Giraudeau (2009) und Campbell (2010) folgende Kriterien zur Beurteilung des Verzerrungsrisikos, um so

Strukturungleichheiten zwischen den Behandlungsgruppen in den eingeschlossenen Clustern und Patienten zum Zeitpunkt der Randomisierung und der Erfassung der Endpunkte zu erkennen:

- Die Beurteilung von Rekrutierungsbias (engl. „recruitment bias“) in der verdeckten Therapiezuweisung beurteilten wir auf zwei Hierarchieebenen, der Ebene der Cluster- (Arztpraxen, Gesundheitszentren) und der der Patienten. Auf der Patientenebene kontrollierten wir, ob entweder alle oder eine Zufallsstichprobe der Patienten aller Cluster in die Studie eingeschlossen wurden und beurteilten dann das Verzerrungsrisiko für die verdeckte Therapiezuweisung als „gering“.
- Ein möglicher Verlustbias (engl. „attrition bias“) muss ebenfalls auf beiden Hierarchieebenen beurteilt werden. Alle Cluster und Patienten sollten in die Auswertung einbezogen werden, die Anzahl der Studienabbrecher sollte gering sein und Gründe für Studienabbrüche gruppenweise angegeben werden. Bei leeren Clustern oder unterschiedlichen Verlustraten in den Randomisierungsgruppen beurteilten wir das Risiko eines Verlustbias mit „hoch“.
- Weitere, daraus resultierende Verzerrungsquellen sehen wir in Unterschieden in der Verteilung prognostisch wichtiger Faktoren zwischen den Behandlungsgruppen zu Studienbeginn und ungeeigneten Methoden in der Datensynthese mit fehlenden Adjustierungen oder der Berücksichtigung hierarchischer Strukturen.

Zusätzlich erfassten wir folgende potentielle Verzerrungsquellen (Unverzagt 2011 und 2015):

- Ungleichheiten in der Verteilung prognostisch wichtiger Faktoren zwischen den Therapiegruppen zu Studienbeginn,
- die Häufigkeit von und der Umgang mit Patienten, welche die Therapie ihrer randomisierten Behandlungszuweisung vor der Endpunkterfassung wechselten (engl. „cross-over“),
- vorzeitige Studienabbrüche auf der Grundlage von Zwischenauswertungen ohne entsprechende Planung und Anpassung des alpha-Fehlers und
- ausschließliche Veröffentlichung von Ergebnissen der per-Protokoll-Auswertungen.

Auch die Beurteilung von Nebenwirkungen kann verzerrt werden, wenn Patienten aus diesen Analysen ausgeschlossen wurden, es keinen klaren Vorgaben zu deren Definition und zum Monitoring gibt oder Nebenwirkungen nicht mit gruppenweisen Häufigkeitsangaben berichtet werden (Cochrane Collaboration 2011, Kapitel 14.6.1). Wir bewerteten das Verzerrungsrisiko des beobachteten Behandlungseffektes je Kriterium auf Grundlage der veröffentlichten Informationen und Nachfragen bei den Autoren als „gering“, „hoch“ oder „unklar“.

Aufgrund verschiedener, in Unverzagt (2011) beobachteter Studieneigenschaften planten wir eine systematische Untersuchung des Einflusses verschiedener Verzerrungsrisiken auf die gemessenen Behandlungseffekte (Unverzagt 2013) auf der Grundlage von Metaanalysen aus der Intensiv- und Notfallmedizin. Datengrundlage dieser meta-epidemiologischen Arbeit waren in der Cochrane Library (2011; 1) veröffentlichte SRs zu den Indikationen kardiogener oder septischer Schock und schwere Sepsis. Dazu suchten wir nach „shock“ oder „sepsis“ im Titel, in den Schlagwörtern (engl. „keywords“) oder im Abstract. Unter diesen Indikationen leidet ein hoher Anteil der auf der Intensivstation behandelten Patienten (Annane 2005, Moerer 2009, Thom 2006).

Der Behandlungseffekt aller untersuchten Interventionen maßen wir an Unterschieden in der Gesamtmortalität zwischen den Behandlungsgruppen, wobei ein Odds Ratio (OR) kleiner 1 einen Behandlungsvorteil in der Interventionsgruppe beschreibt. Zwei Autoren wählten unabhängig voneinander alle Übersichtsarbeiten und Metaanalysen aus, welche mindestens drei RCTs mit dem Endpunkt Mortalität einschlossen. Anschließend entwickelten wir auf der Grundlage des Cochrane Handbuchs (Cochrane Collaboration 2011, Kapitel 8.5.a) und weiterer meta-epidemiologischer Arbeiten (z.B. Balk 2002, Fergusson 2002, Gluud 2006, Hartling 2009, Rosén 2009, Tierney 2005, van Niewenhoven 2001) eine Liste von Vorschlägen möglicher Verzerrungsquellen und Kriterien zu deren Bewertung. Diese Verzerrungsquellen und deren Bewertungskriterien legten wir nach der Bewertung von 16 RCTs in einer Pilotstudie abschließend fest (Unverzagt 2013, Table 1).

Zwei unabhängige Gutachter extrahierten, bewerteten und diskutierten Unterschiede in Studieneigenschaften wie der Anzahl randomisierter und verstorbener Studienteilnehmer je Behandlungsgruppe, der mono- oder multizentrische Durchführung der

Studie und dem Verzerrungsrisikos anhand der elf in Tab. 1 in Unverzagt 2013 definierten Studieneigenschaften (genauere Angaben in Unverzagt 2013). Aus der Anzahl der in den Behandlungsgruppen verstorbenen Patienten generierten wir anschließend individuelle Patientendaten. Wir prognostizierten den Behandlungseffekt in Bezug auf das Überleben der individuellen Patienten in Abhängigkeit von den Bewertungen von zwölf binären Studieneigenschaften.

Dazu modeliierten wir in einem logistischen Modell den Einfluss dieser Studieneigenschaften auf den Behandlungserfolg p_{ij} (Siersma 2007, Unverzagt 2013):

$$\log it(p_{ij}) = b_o + b_{treat} I_t + \sum_{k=1}^{12} (\beta_k c_k(i) I_t + \gamma_k c_k(i)) + \sum_{s=1}^{11} (\delta_s I_s + \eta_s I_s I_t) + \lambda_i I_i$$

Die Indikatoren I_t, I_s, I_i und $c_k(i)$ beschreiben die Behandlung des Patienten j in der Interventions- (t=1) oder Kontrollgruppe (t=0) der Studie i, in welche der Patient behandelt wurde (I_i, i=1,…,82), die Metaanalyse, in welche die Studie eingeschlossen wurde (I_s, s=1,…, 12) und die Studieneigenschaften je Studie c_k (i) (k=1,…12, i=1,…,82). Für diese erste Analyse wählten wir ein Fixed-effects-Modell (FEM) und rechneten dieses mit der SAS-Prozedur PROC LOGISTICS (adaptiert nach Vorschlägen von Kuss 2002, Sterne 2000 und Siersma 2007). Es wurden die Regressionskoeffizienten für den Behandlungseffekt (b_{treat}), 12 Interaktionen zwischen Bewertung und Behandlung (β_k), 12 Studieneigenschaften (γ_k), 81 Studien (λ_i) (im Vergleich zur letzten Studie), 11 Metaanalysen (δ_s) und Interaktionen zwischen Behandlung und Metaanalyse (η_s) (im Vergleich zur letzten Metaanalyse) auf der Basis von 24657 Beobachtungen an individuellen Patienten geschätzt. Zusätzlich untersuchten wir in getrennten unimodalen Analysen den Einfluss jeder Studieneigenschaft.

Die berechneten Ratio of Odds Ratios (ROR) vergleichen die ORs von Studien für einen positiven Behandlungseffekt (OR<1) mit hohem oder unklarem mit denen von Studien mit geringem Verzerrungsrisiko und von mono- mit multizentrischen Studien. Sie quantifizieren so die Modifikation des Behandlungseffektes durch die untersuchten Studieneigenschaften:

$$ROR = \frac{OR_{high\ or\ unclear\ risk\ of\ bias}}{OR_{low\ risk\ of\ bias}} = e^{\beta_k} \quad \text{für k=1,…,11}$$

$$\text{und } ROR = \frac{OR_{monozentri\ sch}}{OR_{multizentr\ isch}} = e^{\beta_k} \text{ für k=12}$$

Ein ROR<1 beschreibt eine Überschätzung des Behandlungseffektes in Studien mit hohem Verzerrungsrisiko oder in monozentrisch durchgeführten Studien.

Anschließend erweiterten wir das Modell zu einem logistischen random effects Modell (REM) und modellierten dieses nach einem Vorschlag aus Kuss (2002) wie folgt:

$$\log it\,(p_{ij} \mid u_i) = b_o + b_{treat} I_t + \sum_{k=1}^{12} (\beta_k c_k(i) I_t + \gamma_k c_k(i)) + \sum_{s=1}^{11} (\delta_s I_s + \eta_s I_s I_t) + u_i$$

Dabei beschreiben I_t, I_s, I_i und $c_k(i)$ die Behandlung des Patienten j in der Interventions- (t=1) oder Kontrollgruppe (t=0) (I_t, t=1-2) in der Studie i, in welcher der Patient behandelt wurde (I_i, i=1,…,82), die Metaanalyse, in welche die Studie eingeschlossen wurde (I_s, s=1,…, 12) und die Studieneigenschaften (k=1,…12, i=1,…,82). Der Parameter u_i ist normalverteilt und erlaubt eine zufallsbedingte Variation der Regressionskoeffizienten und ihrer Standardfehler. Dieses Modell betrachteten wir als realistischer und nutzten es deshalb als Hauptanalyse (SAS, PROC NLMIXED, adaptiert nach Kuss 2002), da es den Einfluss nicht berücksichtigter Studienparameter auf den Behandlungseffekt berücksichtigen kann. Es nutzt die Regressionskoeffizienten des FEM als Anfangswerte. Die verbleibende, nicht erklärte Varianz beschreibt die durch das Modell nicht erklärte Heterogenität (Higgins 2003).

3.2.4 Schätzung des Behandlungseffektes in den Einzelstudien

Die Schätzung des Behandlungseffektes der IABP basierte auf Hazard-Raten (HRs) und deren 95 % KI, soweit diese in der Veröffentlichung berichtet oder wir sie aus individuellen Patientendaten nachberechnen konnten. HRs kleiner 1 beschreiben eine geringere Mortalität in der Interventionsgruppe. Bei fehlenden HRs oder von Informationen, aus denen Standardfehler ermittelt werden können, kontaktierten wir die Erstautoren der Studien oder berechneten diese Parameter aus den veröffentlichten Informationen. Dazu stehen verschiedene direkte und indirekte Methoden zur Verfügung (Parmar 1998, Tierney 2007, Williamson 2002).

Für den Vergleich von Überlebensraten zu festen Zeitpunkten berechneten wir ORs und deren 95 % KI. ORs kleiner 1 beschreiben eine geringere Mortalität in der Interventionsgruppe.

Die Wirksamkeit von Implementierungsstrategien beurteilten wir anhand von Vergleichen der LL-Konformität der Ärzte zwischen den Behandlungsgruppen und nutzte dazu ORs (Unverzagt 2014b, 2014c). ORs über 1 stehen für einen positiven Effekt der Im-

plementierungsstrategien mit höherer Arztadhärenz in der Interventionsgruppe. Die ORs basieren auf dichotomen und metrischen Messskalen aus den Angaben der Ärzte oder Patientenakten. Gegebene ORs aus hierarchischen Modellen mit Berücksichtigung der Clusterstruktur übernahmen wir aus den Einzelstudien. Beschrieben mehrere ORs die Übereinstimmung des Handelns des Arztes zu LL-Empfehlungen, ermittelten wir gewichtete Mittelwerte über alle logarithmierten ORs und ihre Standardfehler. Waren keine ORs gegeben, schätzten wir diese und deren Standardfehler aus den berichteten relativen Risiken und dem Risiko der Kontrollgruppe (Cochrane Collaboration 2011, Kapitel 9.2.2), den angegebenen Häufigkeitsangaben oder der standardisierten Mittelwertdifferenz und deren Varianz (Borenstein 2009). Standardfehler ermittelten wir aus den berichteten Konfidenzintervallen (Borenstein 2009). Ergebnisse aus clusterrandomisierten Studien ohne Berücksichtigung hierarchischer Strukturen korrigierten wir mit dem berichteten Intraclusterkorrelations- (ICC-) Koeffizienten und der mittleren Anzahl von Patienten je Cluster (Cochrane Collaboration 2011, Kapitel 16.3.4). Fehlende ICC-Koeffizienten basierten auf einem medianen ICC-Koeffizient (Campbell 2005) von 0,05 für Studien mit prozessorientiertem Endpunkten in der Allgemeinmedizin.

Wir entschieden uns für die Berechnung relativer Effektgrößen, da sich die Probanden der Einzelstudien in beiden Übersichtsarbeiten sehr stark in ihrem Prä-Interventionszustand unterschieden und diese Unterschiede absolute deutlich stärker als relative Effektmaße beeinflussen (Deeks 2002).

3.2.5 Zusammenfassung von Behandlungseffekten in Metaanalysen

Die Verwendung aggregierter Daten schließt die Kontrolle und Aktualisierung der verwendeten Daten aus. Zusätzliche Analysen oder ein einheitlicher Umgang mit Studienabbrechern über alle eingeschlossenen Einzelstudien hinweg sind nicht möglich. Eine „Metaanalyse mit individuellen Patientendaten“ (IPD, engl. „individual patient data“) setzt voraus, dass die Daten jedes einzelnen Patienten bekannt und die Studie in die Metaanalyse als Effekt und nicht als Beobachtungseinheit eingeht. IPD bieten im Gegensatz zur Verwendung veröffentlichter, aggregierter Daten die Möglichkeit, Metaanalysen anhand der neuesten Daten zu rechnen. Dies ist gerade in der Analyse von Überlebenszeiten mit bei Studienschluss häufig eingeschränkten Nachbeobachtungszeiten sinnvoll, wenn viele Patienten bei Studienende noch leben und ihre Überlebens-

zeiten deshalb zum Zeitpunkt der publizierten Auswertung zensiert wurden. Auch die Aufdeckung und Verringerung methodischer und klinischer Heterogenitätsquellen, die entstehen, wenn die eingeschlossenen Studien sich im Studiendesign, den angewandten Methoden zur Datenauswertung, in den angewandten klinischen Prozeduren oder in Eigenschaften der Studienteilnehmer unterscheiden und dadurch Unterschiede in den beobachteten Behandlungseffekten entstehen (Piedbois 2004), erfordert eine Analyse der IPD. Subgruppenanalysen und Metaregressionsanalysen auf der Grundlage von IPD können dann helfen, die beobachtete Heterogenität sowohl innerhalb als auch zwischen den Einzelstudien zu erklären (Smith 2005). Auch die Eignung von Surrogatendpunkten (hämodynamische Parameter, Arzt- oder Patientenadhärenz) und ihre Korrelation mit patientenrelevanten Endpunkten wie dem Gesamtüberleben kann bei vorliegenden IPD nachgewiesen werden (z.B. Burzykowski 2008).

Aus diesen Gründen planten wir den SR zur Wirksamkeit und Sicherheit der IABP auf der Grundlage von IPD. In den eingeschlossenen Studien lag die Entscheidung zur Weitergabe der Patientendaten bei den Studienleitern und Erstautoren der Studienberichte, obwohl einige dieser Studien von den Herstellern der untersuchten Unterstützungssysteme (Cardiac Assist, Datascope, Abiomed Europe) unterstützt wurden. Wir konnten für einen großen Anteil der eingeschlossenen Einzelstudien in Unverzagt (2015) IPD erhalten. Im Gegensatz dazu lag die Entscheidung zur Weitergabe der IPD in anderen, nicht in diese Arbeit einbezogenen Übersichtsarbeiten (Wagner 2009, 2012) zur Wirksamkeit und Sicherheit anti-angiogenetischer Therapien bei den Sponsoren der Studien (Genentech und Roche). Diese stellten ihre Daten trotz intensiver Bemühungen nicht zur Auswertung zur Verfügung.

Unser Vorgehen zur Initiierung der Zusammenarbeit, der Datensammlung und Validierung von IPD basierte auf Hinweisen von Stewart & Clarke (1995). Nach Identifizierung der relevanten Studien kontaktierten wir alle Erstautoren mit einer Einladung zur Zusammenarbeit in der Zusammenführung, gemeinsamen Auswertung und Diskussion der erhobenen Studiendaten. Diese Einladung enthielt eine Beschreibung der Ziele der geplanten Übersichtsarbeit, der benötigten Daten und möglicher Datenformate, der geplanten Analysen und Informationen zur Datenspeicherung. Nach wiederholten telefonischen und elektronischen Kontakten wurden uns die erforderlichen Daten übermittelt, wir konnten diese kontrollieren und in eine gemeinsame Datenbank einlesen.

Alle Autoren, welche Daten zur Verfügung stellten, waren als Autoren an der Übersichtsarbeit (Unverzagt 2011) beteiligt, kontrollierten die Darstellung der Ergebnisse und hatten die Gelegenheit, die Diskussion und Schlussfolgerungen der Arbeit zu kommentieren.

Die Metaanalysen in Unverzagt (2011) und (2015) basieren auf einem stratifizierten Cox-Modell, in welchem die Behandlungseffekte innerhalb einer Studie berechnet wurden und anschließend mit unterschiedlichen Baseline-Hazard-Funktionen in die Schätzung des Gesamt-HRs eingingen (Whitehead 2002). Dieses Modell wird im Folgenden als Einschritt-Modell bezeichnet. Zusätzlich rechneten wir Cox-Analysen mit Adjustierungen für prognostisch interessante Kovariablen wie Alter, Geschlecht und dem Auftreten von Diabetes sowie Subgruppenanalysen für diese drei Merkmale. Alle diese Analysen basieren auf IPD. In einem Zweischritt-Modell fassten wir die aggregierten Behandlungseffekte aller Einzelstudien zusammen (Riley 2007, 2010) und konnten so Studien mit und ohne IPD gemeinsam auswerten. Aufgrund der hohen Heterogenität zwischen den eingeschlossenen Studien wählten wir in allen Modellen ein REM.

Im Gegensatz dazu basieren die Metaanalysen in Unverzagt (2014b) und (2014c) auf aggregierten Effektschätzern und einer Datensynthese mit einem REM. Das REM wählten wir in beiden Übersichtsarbeiten, da wir bereits in der Planungsphase davon ausgingen, dass sich die in den Einzelstudien beobachteten Behandlungseffekte nicht nur zufällig voneinander unterscheiden. Diese Unterschiede der Studieneigenschaften betrafen die Zusammensetzung der Probanden, die Studienumgebung, den Einsatz der Interventionen, Vergleichsgruppen und die genaue Festlegung der Endpunkte.

Zur Zusammenfassung der Effektschätzer verwendeten wir für HRs, adjustierte ORs und kummulative Inzidenzen die inverse-distance Methode und für dichotome Daten die Mantel-Haenszel-Methode. Zur Berechnung nutzten wir RevMan (Review Manager 2014) oder SAS und Vorschläge von van Houwelingen (2002).

3.2.6 Heterogenitätsanalysen

Die statistische Heterogenität quantifizierten wir in allen Übersichtsarbeiten über den Anteil der Varianz der Behandlungseffekte aufgrund von Studiendifferenzen (I^2-Wert) (Higgins 2003). Die Interpretation folgt dem Cochrane Handbuch (The Cochrane Collaboration 2011, Kapitel 9.5.2). Diese schlagen die Interpretation eines I^2 –Wert von weniger als 40 % als „gering“, von 30 bis 60 % als „moderat“, 50 bis 90 % als „beträcht-

lich“ und 75 bis 100 % als „erheblich“ vor. Die überlappenden Bereiche basieren auf der Abhängigkeit von I^2 von der Studiengröße der Einzelstudien, spiegeln die Unsicherheit in der Bewertung wider und erlauben so einen gewissen Interpretationsspielraum. Zusätzlich beurteilten wir die statistische Heterogenität an der Varianz der Studieneffekte (τ^2) und rechneten Chi-Quadrat-Tests. Diese prüfen die Nullhypothese, dass allen Studien der Metaanalyse ein Behandlungseffekt in derselben Größenordnung zugrunde liegt. Bei hoher statistischer Heterogenität berechneten wir spezifische Effektschätzer für die in der Protokollphase festgelegten potentiellen Effektmodifizierer. Alle diese die Heterogenität beschreibenden Parameter können mit dem Programm RevMan ermittelt werden (Review Manager 2014).

Die Liste der Studieneigenschaften, welche zu unterschiedlich großen Effekten führen können, mussten wir in Unverzagt (2014c) nach dem Lesen der Studien, aber vor der Datenextraktion aufgrund der Diversität der einbezogenen Studien in Hinblick auf die eingeschlossenen Patienten (z.B. Krankheitsschwere), der Zielgruppe der Implementierungsstrategie und von Unterschieden im Studiendesign ergänzen. Wir berichten den Einfluss aller untersuchten Studieneigenschaften. Von den im Protokoll festgelegten acht möglichen Implementierungsstrategien (Shojania 2004) konnten wir sieben in die Meta-Regressionsanalyse einschließen. Eine Strategie (Unterstützung von Datenflüssen) wurde in nur einer der eingeschlossenen Studien untersucht und deshalb aus der statistischen Analyse ausgeschlossen. Die auf Grundlage des Volltextscreenings festgelegten Studieneigenschaften, welche zu einer Effektmodifikation führen können, beinhalteten klinische Unterschiede in den PICO-Kriterien (das für die LL-Implementierung verantwortliche medizinische Personal, Patienten in der Primär-, Sekundär- und Tertiärprävention von CVD und die Länge der Nachbeobachtungszeit) und methodische Unterschiede (Definition des Hauptzielkriteriums in den Einzelstudien, die individuelle Randomisierung von Patienten oder der Cluster-Randomisierung von Ärzten, Praxen oder Kliniken und die Bewertung des potentiellen Verzerrungsrisikos).

In einem Mehrebenen-Modell untersuchten wir den Behandlungseffekt der dichotomisierten uni- oder multimodalen Implementierungsstrategien im Vergleich zu passiven Strategien (=usual care) und den Einfluss von jeweils einem Effektmodifizierer. Wir bezogen alle Vergleiche zu passiven Strategien in das Modell ein, so dass teilweise mehrere Vergleiche je Studie möglich waren. Die log-transformierten ORs waren hin-

reichend normalverteilt und ein gewichtetes lineares gemischtes REM-Modell wurde angepasst. Die Gewichtung folgte aus der Präzision der Schätzungen für das OR der einzelnen Studien mit der inverse-distance Methode.

$$\ln(OR_{ij}) = b_o + \sum_{k=1}^{7} \beta_k I_k + \delta_s I_s + u_{ij}$$

Dabei beschreibt OR_{ij} das OR der Arzt-Adhärenz für den j-ten Vergleich in der i-ten Studie (i=1,…,75; j=1,..,l; l=1,…,3), I_k sind Indikatoren für alle Vergleich von aktiven und passiven Implementierungsstrategien (I_k, k=1,…,7) und I_s beschreiben die Effektmodifikation durch dichotomisierte Studieneigenschaften (I_s, s=1,…, 6). Der Parameter u_{ij} ist normalverteilt und erlaubt eine zufallsbedingte Variation der in den Studien und Vergleichen geschätzten ORs. Diese Analysen rechneten wir mit der SAS-Prozedur PROC MIXED. Zusätzlich zur Untersuchung des prognostischen Einflusses der sieben Implementierungsstrategien untersuchten wir die Effektmodifikation durch jeweils eine der sechs Studieneigenschaften.

Die Effektmodifikation durch die klinischen und methodischen Studieneigenschaften beschreiben wir über RORs und deren 95 % KI. Diese quantifizieren den Faktor, um welche die ORs der Arzt-Adhärenz durch die spezifischen Studieneigenschaften verändert werden kann. Konfidenzintervalle, welche die 1 nicht enthalten, gelten als signifikant. Das Ausmaß der Effektmodifikation durch die untersuchten Studieneigenschaften bewerten wir am Anteil der Varianz zwischen den Studien (τ^2), welcher durch sie erklärt werden konnte.

4. Ergebnisse

4.1 Wirksamkeit und Sicherheit der IABP im kardiogenen Schock: von der Evidenzgenerierung zu veränderten Leitlinienempfehlungen

4.1.1 IABP SHOCK-Studie

Insgesamt 45 Patienten mit infarktbedingtem kardiogenen Schock wurden in die Studie aufgenommen, von denen 23 dem Behandlungsarm mit IABP und 22 dem Standardtherapiearm zugewiesen wurden. Vier Patienten entsprachen nicht den Einschlusskriterien, bei einem weiteren Patienten wurden keine studienspezifischen Messungen erhoben oder studienkonforme Therapiemaßnahmen eingeleitet, so dass die Auswertung auf insgesamt 40 randomisierten Patienten basiert. Eine Nachrekrutierung der fünf ausgeschlossenen Patienten war nach einer Ergänzung des Studienprotokolls und der Zustimmung der zuständigen Ethikkommission möglich und erfolgte unabhängig von den vorliegenden Daten.

Einer der 21 Patienten, welche dem Standardtherapie-Arm zugewiesen wurden, wechselte in den IABP-Arm. Von den randomisierten und erfolgreich eingeschlossenen Patienten konnte bei drei Patienten im IABP-Arm (ein Todesfall, zwei Patientenverlegungen von der Intensivstation) sowie zwei Patienten im Standardtherapie-Arm (ein Todesfall, eine Verlegung von der Intensivstation) die Datenerhebung nicht vollständig über den viertägigen Beobachtungszeitraum vorgenommen werden (Figure 1 in Prondzinsky 2010).

Wirksamkeit und Sicherheit

Im IABP-Arm wurde während der viertägigen Beobachtungsperiode ein mittlerer Abfall des Apache-II-Scores um 4.4±6.8 Punkte im Vergleich zu 3.3±5.9 Punkten im Standardtherapiearm beobachtet. Der geringe mittlere Unterschied von 1.09 (95 % KI:-3.01 bis 5.19) Punkten war weder klinisch relevant noch statistisch signifikant (p = 0.591). Die Krankenhausmortalität betrug 36.8 % (7 Verstorbene von 19 Patienten) im IABP-Arm und 28.6 % (6 Verstorbene von 21 Patienten) im Standardtherapie-Arm.

Auch eine Verbesserung der hämodynamischen, pro- und anti-inflammatorischen Parametern durch den zusätzlichen Einsatz der IABP konnte nicht nachgewiesen werden, während der vermutete prognostische Wert aller erhobenen Parameter bestätigt werden konnte (Prondzinsky 2012a, 2012b, 2012c). Eine möglicherweise durch die IABP

hervorgerufene, nicht lebensgefährdende Komplikationen (Beinischämie) wurde bei einem Patienten in der Behandlungsgruppe festgestellt.

Die Studienergebnisse reichten wir ab 2007 bei mehreren hochrangigen kardiologischen Zeitschriften zum Reviewverfahren ein. Sie wurden nach zwei Ablehnungen und ausführlichen Sensitivitäts- und Subgruppenanalysen schließlich fünf Jahre nach Studienabschluss publiziert (Prondzinsky 2010). Zeitgleich wurden mehrere RCT zur Wirksamkeit der IABP bei Patienten mit infarktbedingtem kardiogenen Schock durchgeführt, deren Ergebnisse in einem SR zusammengefasst wurden.

4.1.2 Systematische Übersichtsarbeit

Aus insgesamt 1410 Referenzen aus der systematischen Suche konnten wir im Jahr 2010 sechs abgeschlossene und zwei laufende geeignete Studien identifizieren. Zwei Studien rekrutierten zu diesem Zeitpunkt noch Patienten. Eine Studie wurde 2012 publiziert (IABP Shock II: Thiele 2012), während die andere nach Einschluss eines Patienten wegen Rekrutierungsschwierigkeiten abgebrochen wurde. Die systematische Suche für den SR aktualisierten wir im Januar 2013. Im Rahmen dieser systematischen Suche waren erneut 728 Referenzen zu screenen und 15 Volltexte zu lesen. Wir konnten eine zusätzliche Studie in die Arbeit einschließen (IABP Shock II: Thiele 2012) (Abb. 1).

Die aktualisierte Arbeit enthält Daten von 790 Patienten mit Myokardinfarkt und kardiogenem Schock aus sieben Studien und uns stehen individuelle Patientendaten für sechs Studien mit insgesamt 750 Patienten zur Verfügung (Unverzagt 2015).

Vier der eingeschlossenen Studien vergleichen eine Behandlung mit IABP mit einer Behandlung ohne IABP und drei weitere vergleichen eine Behandlung mit einer IABP mit einer Behandlung mit anderen linksventrikulären Unterstützungssystemen wie TandemHeart oder Impella. Insgesamt 406 Patienten wurden in den IABP-Arm und 384 Patienten in Kontrollgruppen randomisiert, von denen 339 ohne und 45 mit anderen Unterstützungssystemen behandelt wurden.

Alle sieben eingeschlossenen Studien berichten die Mortalität für das Kurzzeitüberleben während des Aufenthaltes im Krankenhaus oder über 30 Tage, während Informationen zum Langzeitüberleben für vier Studien vorliegen.

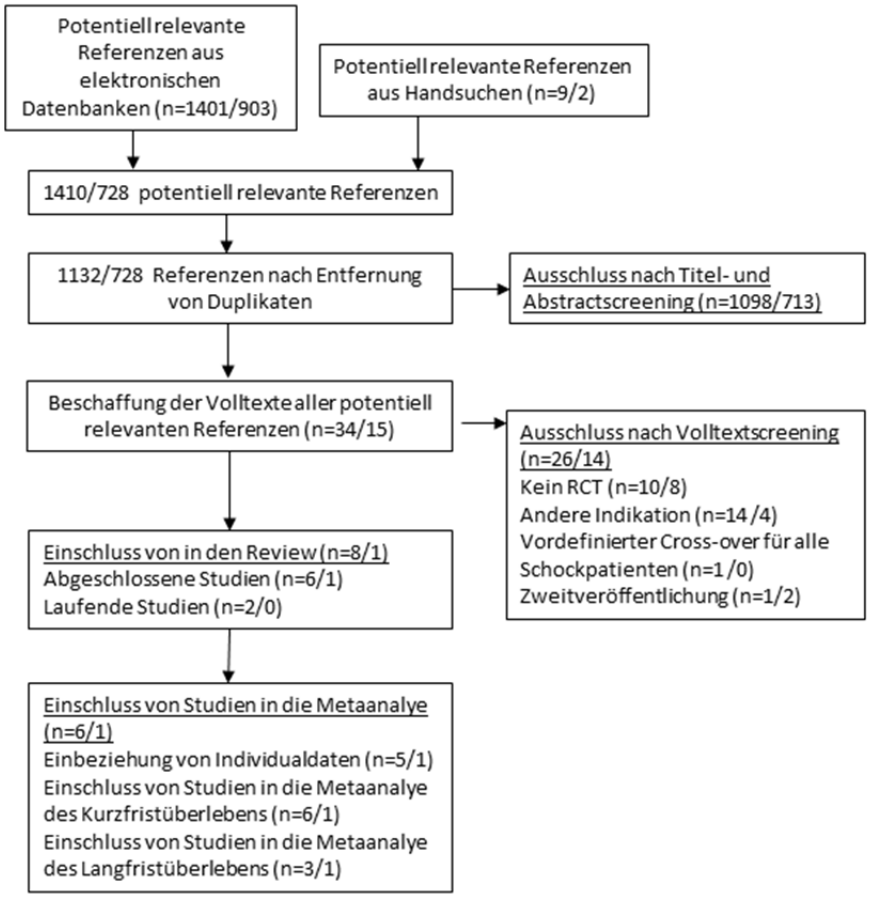


Abb. 1. Flussdiagramm für die systematische 1. / 2. Suche im Januar 2010 und 2013 in der Übersichtsarbeit zur Wirksamkeit der IABP (angepasst aus Unverzagt 2011 und 2015, figure 1).

Die extrahierten Informationen enthalten Informationen (Unverzagt 2015, Characteristics of included studies) zu

- den eingeschlossenen Patienten,
- Behandlungen und hämodynamische Parametern zum Zeitpunkt der Randomisierung,
- Informationen zum Einsatz der IABP,
- Endpunkten und
- der Häufigkeit von Studienabbrechern und deren Ausschlussgründe.

Eine detaillierte Beschreibung der einzelnen Studien, der Studienteilnehmer, untersuchten Endpunkte und eingesetzten Intervention wird in Unverzagt (2015) in Anlage 1 dieser Arbeit gegeben.

Bewertung des Verzerrungsrisikos

Die Verteilung der Verzerrungsquellen in den sieben eingeschlossenen Studien ist in Unverzagt (2015) (Figure 3 und 4) für die Einzelstudien und über relative Häufigkeitsangaben detailliert dargestellt.

Auf Grundlage der veröffentlichten und zusätzlich erfragten Informationen beurteilten wir das Verzerrungsrisiko aufgrund von Selektionsbias in der Generierung der Zufallsfolge und der verdeckten Therapiezuweisung in allen Studien mit „gering". Das größte Risiko auf eine verzerrte Schätzung des Behandlungseffektes entstand durch die Schwierigkeiten in der Verblindung von Ärzten und Pflegenden, so dass Unterschiede in der Pflege und weiteren Behandlung des Patienten und Unterschieden in der Endpunkterfassung nicht ausgeschlossen werden können. Von einem geringen Verzerrungsrisiko kann in der Erfassung der Endpunkte nur in der multizentrischen IABP-Shock-II Studie ausgegangen werden, in welcher alle Endpunkte zur Wirksamkeit und Sicherheit durch ein gegenüber der Therapiezuweisung verblindetes Komitee beurteilt wurde. In einer Studie fehlten die üblicherweise in diesen Studien berichteten Informationen zur Hämodynamik.

Neben den standardisiert nach den Cochrane-Kriterien (Cochrane Collaboration 2011, Kapitel 8.5.a) erhobenen Verzerrungsquellen identifizierten wir in drei Studien weitere systematische Verzerrungen aufgrund einer ausschließlichen Veröffentlichung von Per-Protokollanalysen mit einem hohen Anteil von Therapiewechseln, bei vorzeitigem datengesteuertem Studienabbruch ohne vorherige Planung und bei Einschluss von Patienten, welche bereits zum Zeitpunkt der Randomisierung eine IABP erhalten hatten.

Der Funnelplot (Unverzagt 2015, Figure 5) für die HR des 30 Tages-Überlebens zeigte keinen Hinweis auf einen Publikationsbias, sollte aber auch aufgrund der geringen Anzahl eingeschlossener Studien, von denen nur eine (IABP-Shock-II trial, Thiele 2012) mehr als 40 Patienten einschloss, vorsichtig interpretiert werden. Die Effektschätzer der kleinen Studien waren um den gepoolten Behandlungseffekt, welcher weitgehend durch den Effekt der großen multizentrischen Studie bestimmt wurde, symmetrisch verteilt.

Wirksamkeit und Sicherheit

Das Überleben über 30 Tage wurde mit Ausnahme einer Studie vollständig erhoben, zum Langzeitüberleben über 6-Monate lagen Informationen von vier und nach 12 Mo-

naten von zwei Studien vor. Die gruppenweisen Überlebenszeiten basieren auf IPD für sechs Studien und wurden über Kaplan-Meier-Kurven beschrieben (Unverzagt 2015, Figure 2).

Die Poolung von sechs Studien ergab eine 30-Tage-Mortalität von 40,1 % in der Interventionsgruppe (150 Verstorbene von 375 Patienten) und 40,9 % in der Kontrollgruppe (153 Verstorbene von 375 Patienten). Der Forest-Plot (Abb. 2) weist für keine der eingeschlossenen Studien eine relevante Verringerung der Mortalität in der IABP-Gruppe auf. Die gepoolten Gesamteffekte aus dem primären Einschritt- und dem Zweischritt-Modell (HR 0,95; 95 % KI 0,76-1,19) stimmen überein. Auch die Subgruppenanalysen (IABP vs. Standardbehandlung ohne IABP und IABP vs. andere linksventrikuläre Unterstützungssysteme) zeigen keinen Unterschied im Behandlungseffekt der IABP. Insgesamt ist die statistische Heterogenität zwischen den Effektschätzern gering (I^2=0%, T^2<0,01, p=0,97), was teilweise auf die geringen Studiengrößen in fünf der sechs eingeschlossenen Studien zurückgeführt werden kann.

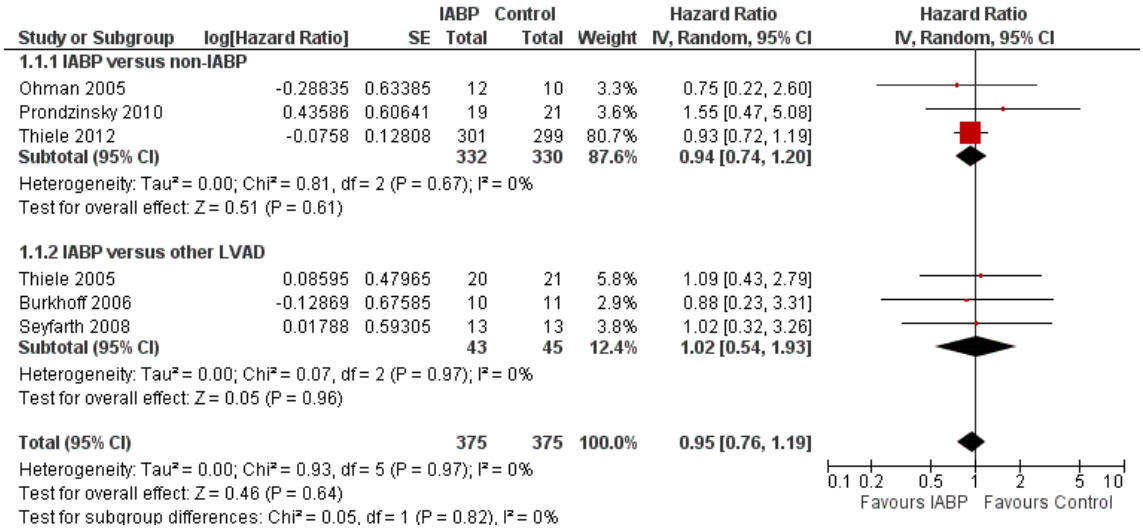


Abb. 2. Forestplot zur Wirksamkeit einer IABP auf das Überleben über 30 Tage (Unverzagt 2015, figure 6).

Die gepoolten Behandlungseffekte sind konsistent und konnten in Subgruppenanalysen für Männer und Frauen, jüngere und ältere Patienten (≥ 75 Jahre) und Patienten mit und ohne Diabetes bestätigt werden. Sensitivitätsanalysen untersuchten den Effektschätzer nach einer Berechnung mit dem Zweischritt-Modell, und verglichen Mortalitätsraten zu festen Zeitpunkten (bis zur Entlassung aus dem Krankenhaus, über 30 Tage, 6 Monate und 1 Jahr). Alle Analysen bestätigen den fehlenden Behandlungsef-

fekt eines IABP-Einsatzes bei Patienten mit infarktbedingtem kardiogenem Schock (Unverzagt 2011, Update 2015) (Additional tables, table 1).

Einige der kleineren Studien beobachteten verbesserte hämodynamische Effekte unter IABP, während Nebenwirkungen wie moderate oder starke Blutungen und Infektionen vorrangig in den Kontrollgruppen mit anderen linksventrikulären aggressiveren Unterstützungssystemen auftraten.

4.2 Einfluss von Studieneigenschaften auf den Behandlungseffekt von Studien der Intensiv- und Notfallmedizin

Sechs SRs mit zwölf Metaanalysen zur Gesamtmortalität, in welchen die Behandlungseffekte von 82 RCTs gepoolt wurden, konnten wir in diese meta-epidemiologische Studie einschließen. Table 2 in Unverzagt (2013) gibt für die in den Übersichtsarbeiten eingeschlossenen Metaanalysen eine detaillierte Beschreibung der Indikationen der Probanden, untersuchten Interventionen, Kontrollgruppen, Patientenzahlen und der Effektschätzer in den Metaanalysen.

Die eingeschlossenen Studien randomisierten zwischen 10 und 2634 Patienten und wurden jeweils zur Hälfte monozentrisch mit einer medianen Studiengröße von 40 Patienten (Range 10 bis 252) und multizentrisch mit einer medianen Größe von 223 Patienten (21 bis 2634) durchgeführt.

Die Ergebnisse der Beurteilung des Verzerrungsrisikos für elf Studieneigenschaften nach den Vorgaben von Table 1 in Unverzagt (2013) zeigen, dass monozentrische im Vergleich zu multizentrisch durchgeführten Studien in nahezu allen Studieneigenschaften häufiger ein hohes oder unklares Verzerrungsrisiko aufwiesen (Tab. 1).

Die Qualität der Bewertung basiert auf den in den Studien oder SRs veröffentlichten Informationen. In den Jahren vor der Veröffentlichung der CONSORT-Richtlinien (Moher 2001), in welchen 58 % der eingeschlossenen Studien publiziert wurden, fehlten häufig für die Beurteilung wichtige Informationen. Deshalb konnten wir einige Studieneigenschaften nicht sicher hinsichtlich ihres Verzerrungsrisikos beurteilen. Wir beurteilten das Verzerrungsrisiko in der verdeckten Therapiezuweisung, der Beschreibung von Studienabbrüchen und der Generierung der Zufallsfolge in 48, 38 und 39 % der Studien mit „unklar“. Eine verzerrte Bewertung der untersuchten Assoziation zum Behandlungseffekt können wir für diese Studieneigenschaften nicht ausschließen.

Tab. 1. Häufigkeit der Bewertung des Verzerrungspotentials mit „hoch“ oder „unklar“ in mono- und multizentrisch durchgeführten Studien.

Studieneigenschaft	Häufigkeit von hohem /unklarem Verzerrungspotential (n (%)) in Monozentrischen Studien (n=41)	Multizentrischen Studien (n=41)
Generierung der Zufallsfolge	2 (4,8 %) / 20 (48,8 %)	1 (2,4 %) / 12 (29,3 %)
Verdeckte Therapiezuweisung	3 (7,3 %) / 29 (70,7 %)	1 (2,4 %) / 11 (26,8 %)
Doppelte Verblindung	20 (48,8 %) / 4 (9,8 %)	10 (24,4 %) / 0 (0 %)
Beschreibung von Studienabbrüchen	5 (12,2 %) / 0 (0 %)	5 (12,2 %) / 1 (2,4 %)
Selektives Berichten der Endpunkte	5 (12,2 %) / 9 (21,9 %)	1 (2,4 %) / 4 (9,8 %)
Früher Studienabbruch	8 (19,5 %) / 25 (61,0 %)	12 (29,3 %) / 7 (17,1 %)
Interventionen vor Studienbeginn	16 (39,0 %) / 7 (17,1 %)	7 (17,1 %) / 3 (7,3 %)
Interessenkonflikte	10 (24,4 %) / 9 (21,9 %)	12 (29,3 %) / 3 (7,3 %)
Unterschiede zu Studienbeginn	22 (53,7 %) / 5 (12,2 %)	13 (31,7 %) / 3 (7,3 %)
Cross-over	3 (7,3 %) / 6 (14,6 %)	3 (7,3 %) / 1 (2,4 %)
Ausreichende Nachbeobachtungszeiten	6 (14,6 %) / 14 (34,2 %)	5 (12,2 %) / 0 (0 %)

n: Anzahl

Tab. 2 fasst die RORs für den untersuchten Zusammenhang zwischen zwölf verschiedenen Studieneigenschaften und dem Behandlungseffekt hinsichtlich der Mortalität zusammen. Die Punktschätzer der ROR weisen für die primäre Analyse Werte zwischen 0,64 und 1,13 auf.

Insgesamt konnten wir nur für eine der zwölf untersuchten Eigenschaften einen signifikanten Zusammenhang zum Behandlungseffekt nachweisen. Monozentrisch durchgeführte Studien überschätzten den Behandlungseffekt im Mittel um 36 % (ROR 0,64; 95 % KI 0,47-0,87). Die quantifizierten Effekte unterscheiden sich für diese Studieneigenschaft kaum zwischen den univariaten und multivariaten, REM- und FEM- Modellen. Dies deutet darauf hin, dass es möglicherweise kein wesentliches Confounding für diesen Zusammenhang gibt.

Weitere fünf Studieneigenschaften zeigen einen Trend zu einem Zusammenhang zum geschätzten Behandlungseffekt. So stellten wir fest, dass Studien mit einem erhöhten Risiko der selektiven Auswahl der berichteten Endpunkte den Behandlungseffekt im Mittel um 20 % überschätzen (ROR 0,80; 95 % KI 0,57- 1,12). Überschätzungen des Behandlungseffektes sind ebenfalls beim Einsatz von Interventionen mit ähnlichen Wirkmechanismen vor der Randomisierung (ROR 0,86; 95 % KI 0,73 – 1,00) und bei zum Zeitpunkt der Randomisierung vorhandenen Unterschieden zwischen prognostisch wichtigen Parametern (95 % KI 0,90; 95 % KI 0,75- 1,07) möglich.

Tab. 2. Zusammenhang zwischen den potentiellen Verzerrungsrisiken in 12 Studieneigenschaften und dem Behandlungseffekt: Primäre Analyse und Sensitivitätsanalysen (aus Unverzagt 2013, table 3 ergänzt um Sensitivitätsanalysen).

Studieneigenschaft	**OR in Studien mit hohem oder unklarem Verzerrungsrisiko / OR in Studien mit niedrigem Verzerrungsrisiko (95% Konfidenzintervall)**			
	Primäre Analyse	**Sensitivitätsanalysen**		
Modell	**REM**	**FEM**	**REM**	**FEM**
Anzahl der untersuchten Eigenschaften	**12**	**12**	**1**	**1**
Generierung der Zufallsfolge	0,97 (0,76 - 1,24)	1,17 (0,86 - 1,59)	0,93 (0,80 - 1,09)	0,99 (0,89 - 1,11)
Verdeckte Therapiezuweisung	1,13 (0,94 - 1,19)	0,92 (0,70 - 1,19)	0,93 (0,81 - 1,06)	1,02 (0,93 - 1,13)
Doppelte Verblindung	1,03 (0,80 - 1,32)	1,02 (0,79 - 1,32)	0,84 (0,69 - 1,02)	0,91 (0,76 - 1,09)
Beschreibung von Studienabbrüchen	1,13 (0,89 - 1,42)	1,26 (0,95 - 1,66)	1,19 (0,98 - 1,45)	1,23 (1,02 - 1,49)
Selektives Berichten der Endpunkte	0,80 (0,57 - 1,12)	0,80 (0,57 - 1,13)	0,73(0,54 - 0,98)	0,84 (0,64 - 1,10)
Früher Studienabbruch	1,10 (0,94 - 1,29)	1,03 (0,82 - 1,29)	0,98 (0,86 - 1,11)	1,02 (0,89 - 1,15)
Interventionen vor Studienbeginn	0,86 (0,73 - 1,00)	0,89 (0,66 - 1,21)	0,89 (0,74 - 1,08)	0,95 (0,79 - 1,14)
Interessenkonflikte	1,05 (0,88 - 1,24)	1,07 (0,89 - 1,29)	1,01 (0,87 - 1,15)	1,04 (0,90 - 1,19)
Unterschiede zu Studienbeginn	0,90 (0,75 - 1,07)	0,97 (0,79 - 1,20)	0,91 (0,80 - 1,04)	0,97 (0,85 - 1,11)
Cross-over	0,89 (0,61 - 1,31)	0,88 (0,59 - 1,30)	0,68 (0,49 - 0,96)	0,69 (0,50 - 0,97)
Ausreichende Nachbeobachtungszeiten	1,11 (0,90 - 1,38)	1,24 (0,95 - 1,62)	1,01 (0,84 - 1,21)	0,99 (0,83 - 1,18)
Multi- vs. monozentrische Studie	0,64 (0,47 - 0,87)	0,65 (0,47 - 0,91)	0,64 (0,50 - 0,80)	0,69 (0,55 - 0,86)

OR: Odds Ratio; REM: Random-effects-Modell; FEM: Fixed-effects-Modell

Aber diese Zusammenhänge sollen aufgrund der durch die breiten Konfidenzintervalle wiedergespiegelten geringen Präzision nur sehr vorsichtig interpretiert werden und bedürfen der Bestätigung auf einer breiteren Studiengrundlage.

Zwei Studieneigenschaften wiesen einen Trend zu einer Unterschätzung des Behandlungseffektes bei hohem oder unklarem Verzerrungsrisiko auf: die verdeckte Therapiezuweisung (ROR 1,13; 95 % KI 0,94 – 1,19) und die Beschreibung von Studienabbrüchen (ROR 1,13; 95 % KI 0,89–1,42).

Abb. 3 beschreibt in einem Funnelplot über alle eingeschlossenen Studien und Metaanalysen die bivariate Varianz der beobachteten Behandlungseffekte und ihrer Präzision. Ein möglicher Hinweis auf Publikationsbias ergibt sich aus den fehlenden kleinen Studien mit hohem Standardfehler und einem geschätzten OR größer 1,5, in welchen höhere Mortalitätsraten in der Interventionsgruppe beobachtet wurden.

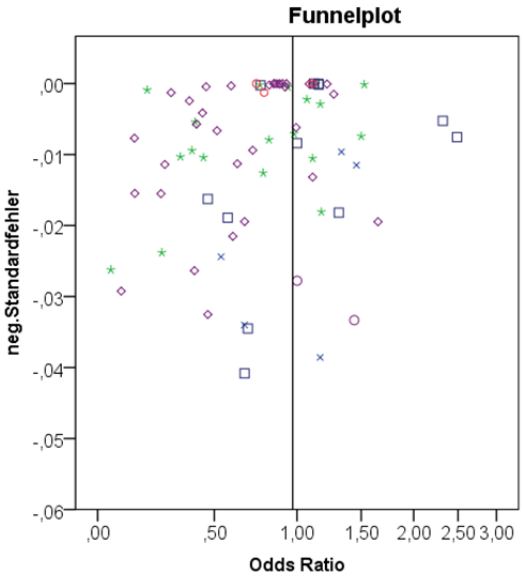


Abb. 3. Abhängigkeit der Präzision (gemessen über den negativen Standardfehlers des logarithmierten ORs) von den geschätzten Behandlungseffekten (ORs in den Einzelstudien). Die sechs verschiedenen Symbole beschreiben die in die Studie eingeschlossenen SRs.

Mit einem Test auf „Small-Study“ Effekte (Egger 1997) überprüften wir, ob die Beobachtungseffekte mit der beobachteten Varianz zunehmen. Diese Vermutung konnten wir nicht bestätigen (Regressionskoeffizient 1,05; 95 % KI 0,96-1,14).

4.3 Wirksamkeit von Implementierungsstrategien auf die leitlinienkonforme Behandlung von Patienten mit Herz-Kreislauferkrankungen

Insgesamt konnten wir über eine systematische Suche 18115 potentiell relevante Referenzen finden. Nach dem Entfernen von Duplikaten bewerteten wir 13384 Referenzen im Titel- und Abstraktscreening und 364 Volltexte im Volltextscreening. Insgesamt 75 Studien mit 84 Vergleichen entsprachen den vordefinierten Einschlusskriterien und wurden in die metaanalytischen Auswertungen eingeschlossen (Unverzagt 2014b,c). Von diesen verglichen 54 Studienarme einer Strategiegruppe zuordenbare (unimodale) Interventionen mit einer passiven Implementierung und 30 Studienarme Kombinationen von mehreren (multimodalen) Interventionen mit einer passiven Implementierung der Leitlinien. Insgesamt ca. 256.500 Patienten mit CVD (z.B. Hypertonie, Hypercholesterinämie, koronare Herzkrankheit) und 8.800 Ärzte, Schwestern und Angehörige anderer Gesundheitsberufe (v.a. Apotheker) wurden in die Studien eingeschlossen. Mehr als ein Drittel der Studien schloss Patienten mit mehreren (bis zu sechs) Indikationen ein.

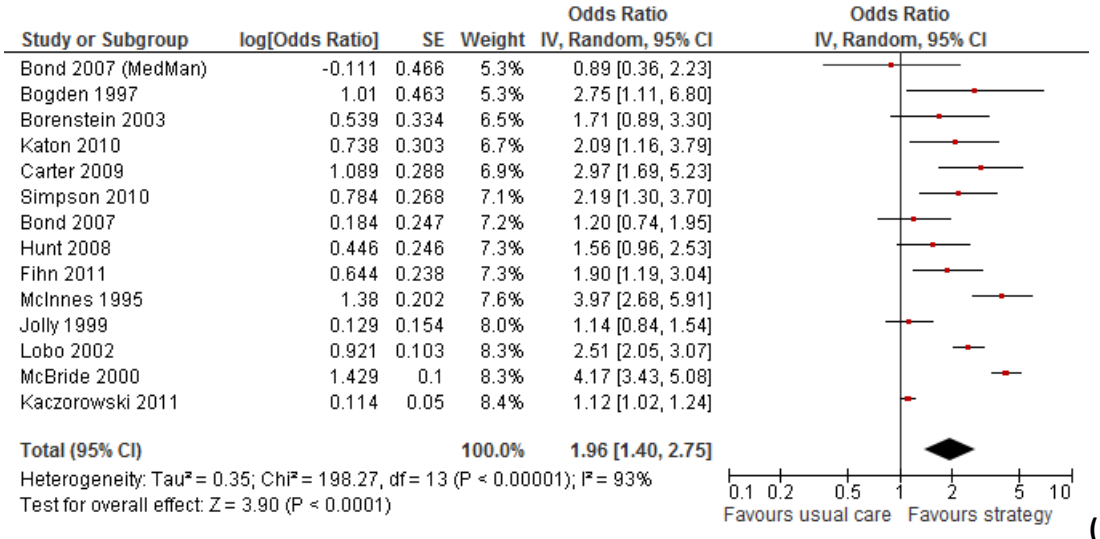
Die Implementierungsstrategien umfassten insgesamt acht Strategiekategorien (Anwenderschulungen, Anwender-Erinnerungssysteme, Unterstützung von Datenflüssen, Audit und Feedback, organisatorische Veränderungen, Patientenschulungen, Patienten-Erinnerungssysteme und Unterstützung des Selbstmanagements der Patienten). Die Kontrollgruppe unterlag anfangs keinen Einschränkungen, die metaanalytischen Auswertungen begrenzten wir dann aber auf Vergleiche zu passiven Implementierungsstrategien, um die hohe Heterogenität zwischen den Studien zu begrenzen.

Der Umsetzungserfolg wurde in allen Studien über den Prozessparameter „Arztadhärenz“ gemessen, welcher über eine Vielzahl von Operationalisierungen wie Medikamentenverordnungen und –dosierungen, dokumentierte Lebensstilinterventionen, Verlaufsbeobachtungen oder den Einsatz spezifischer diagnostischer Maßnahmen, beschrieben worden. Informationen zu patientenorientierten Endpunkten (Lebensqualität, Morbidität, Mortalität) und Kostenaspekten wurden nur in sehr wenigen Studien erhoben und berichtet, so dass keine aussagekräftigen Auswertungen möglich waren.

In Unverzagt (2014b) beschreiben wir detailliert die 54 Studien mit unimodalen Strategien, in denen die Wirksamkeit einer einzelnen Implementierungsstrategie mit der passiven Implementierung verglichen wurde. Die Behandlungseffekte der Einzelstu-

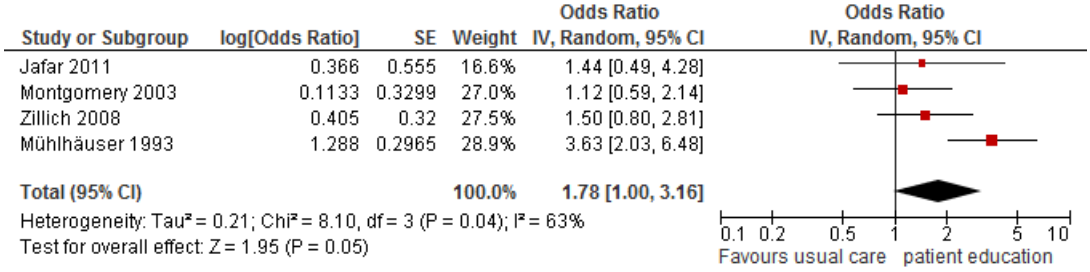
dien visualisieren wir in Forestplots und untersuchten diese in getrennten Metaanalysen auf ihre Wirksamkeit auf eine Verbesserung der Arztadhärenz (Unverzagt 2014b, figure 2 und 3, Abb. 4).

(A) Organisatorische Änderungen im Versorgungsablauf



(B)

(B) Maßnahmen zur Patientenschulung



(C) Einsatz von Anwender-Erinnerungssystemen

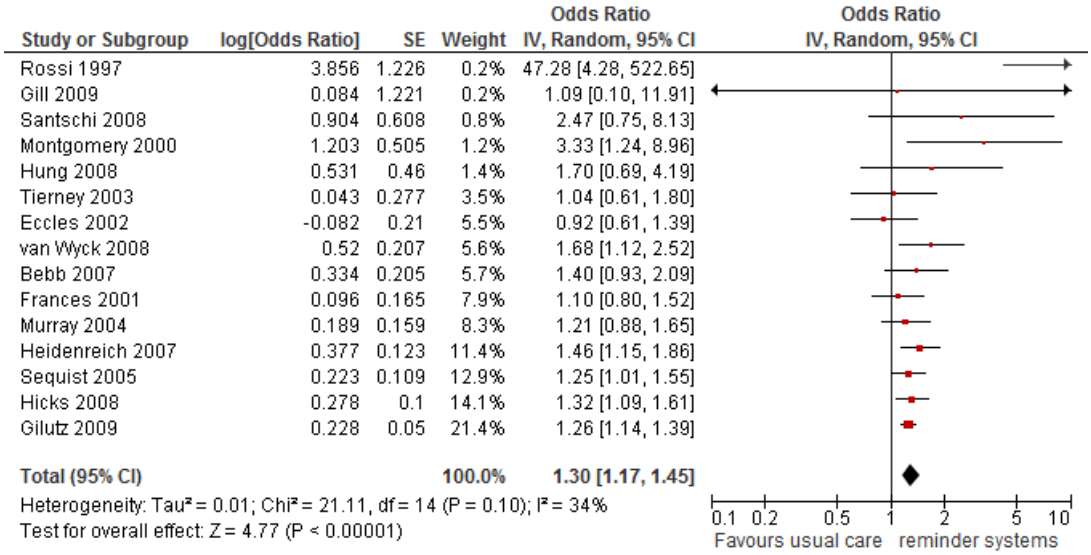


Abb. 4. Zusammenfassung aller Vergleiche von drei erfolgreichen unimodalen Implementierungsstrategien (A,B,C) mit einer passiven LL-Implementierung (aus Unverzagt 2014b, figure 2 und 3 modifiziert).

Diese erste Auswertung bot die Möglichkeit, den Einsatz der sich gegenseitig beeinflussenden Strategien unabhängig voneinander zu beurteilen und damit die Heterogenität zwischen den Studien zu begrenzen. Die Metaanalysen lieferten einen ersten Hinweis auf eine hohe Wirksamkeit von organisatorischen Veränderungen im Versorgungsablauf (OR 1,96; 95 % KI 1,40-2,75). Auch eine Wirksamkeit von Maßnahmen zur Patientenschulung (OR 1,78; 95 % KI 1,00-3.16), Anwenderschulungen (1,40; 95 % KI 1,16-1,68) und des Einsatzes von Anwender-Erinnerungssystemen (OR 1,30; 95 % KI 1,17-1,45), nicht jedoch der übrigen Strategiekategorien konnten wir nachweisen.

Wir empfehlen alle vier genannten Maßnahmen zur Verbesserung der Arztadhärenz, auch wenn die über den I^2-Wert quantifizierte statistische Heterogenität zwischen den Behandlungseffekten von Studien mit organisatorischen Veränderungen im Versorgungsablauf als erheblich (I^2=93 %), beim Einsatz von Maßnahmen zur Arzt- oder Patientenschulung als beträchtlich (I^2=69 % und 63 %) und bei Anwender-Erinnerungssystemen als moderat bezeichnet werden muss (I^2=34 %).

Weitere Analysen zeigten die Wirksamkeit dieser Implementierungsstrategien beim multimodalen Einsatz mehrerer Implementierungsstrategien in insgesamt 75 Studien mit 84 Vergleichen zur passiven LL-Implementierung (Unverzagt 2014c, Table 1). Die Wirksamkeit von organisatorischen Veränderungen im Versorgungsablauf (ROR 1,49; 95 % KI 1,21-1,82) und von Maßnahmen zur Patienten- (ROR 1,48; 95 % KI 1,08-2,01) und Anwenderschulung (1,34; 95 % KI 1,08-1,65) konnten wir im multimodalen Modell bestätigen.

Ein Vergleich der Wirksamkeit multi- und unimodaler Strategien konnten wir aufgrund der geringen Power für keine der untersuchten Implementierungsstrategien einen signifikanten Effekt für die Integration einer oder mehrerer Kategorien in eine multimodale Strategie (-kombination) nachweisen.

Tendenzen für einen klinisch relevanten Zusatznutzen ergaben sich für Audit und Feedback (ROR 1,59; 95 % KI 0,95-2,66) und der Unterstützung des Patientenselbstmanagement (ROR 2,0; 95 % KI 0,97-4,15). Insgesamt stellten wir eine mit der Anzahl der eingesetzten Implementierungsstrategien steigende Arztadhärenz fest (ROR 1,18; 95 % KI 1,09-1,26) (Klement 2015).

4.4. Effektmodifikation beim Einsatz von Implementierungsstrategien

Die auf ihre Effektmodifikation untersuchten klinischen und methodischen Studieneigenschaften waren über die verschiedenen Kategorien der Implementierungsstrategien ungleich verteilt (Tab. 3). LL-Empfehlungen richteten sich in den meisten Studien direkt an in der Primärversorgung tätige Ärzte, welche in 19 (25 %) Studien durch Schwestern oder in 12 (16 %) Studien durch Vertreter anderer Gesundheitsberufe (z.B. durch Apotheker) unterstützt wurden. Die erfolgreichsten Implementierungsstrategien (organisatorische Veränderungen und Maßnahmen zur Patientenschulung) setzten häufig Schwestern und andere Gesundheitsberufe in der medizinischen Versorgung der Patienten ein.

Die meisten Studien (76 %) wurden in der Sekundärprävention von CVD durchgeführt, einige schlossen zusätzlich Patienten in der Primär- und Tertiärprävention (zehn und sechs Studien) ein. Sechs (8 %) Studien konzentrierten sich auf Patienten in der Primärprävention und zwölf Studien (16 %) auf die Tertiärprävention von CVD.

Etwa die Hälfte der Studien war als Prozessoptimierungsstudie mit Arztadhärenz als primärem Endpunkt geplant, die andere Hälfte maß die Wirksamkeit der Implementierungsstrategien an patientenrelevanten Endpunkten. Es wurden in 23 (31 %) Studien Patienten und in 52 (69 %) Studien Ärzte oder Praxen randomisiert. Die Nachbeobachtungszeiten zeigten eine Spannbreite von 3 bis zu 36 Monaten. In 61 Studien lag die Nachbeobachtungszeit bei einem Jahr oder darunter.

In der Bewertung des Verzerrungsrisikos konnten wir nur 15 (20 %) der eingeschlossenen Studien mit einem geringen Verzerrungsrisiko in allen untersuchten Kriterien beurteilen. Ein häufiger Kritikpunkt war die fehlende Festlegung des Hauptzielkriteriums in der Fallzahlanalyse in 28 (37 %) Studien, so wir eine selektive Wahl des Hauptzielkriteriums nicht sicher ausschließen konnten.

Außerdem wurde die Arztadhärenz gerade in clusterrandomisierten Studien, welche häufig zur Verbesserung von patientenrelevanten Endpunkten geplant waren, nicht immer für prognostisch wichtige, aber zu Studienbeginn ungleich verteilte Faktoren adjustiert. Aus diesem Grund bewerteten wir das Verzerrungspotentials in 28 (37 %) Studien mit „hoch“.

Tab. 3. Behandlungseffekte der Implementierungsstrategien auf die Adhärenz der Ärzte (OR) und Häufigkeitsverteilung der Effektmodifizierer je eingesetzter Kategorie von Implementierungsstrategien (modifiziert nach Unverzagt 2014b und c, table 1).

Implementierungs-Strategie	OR (95% KI) unimodales Modell (Anzahl Studien)	ROR (95% KI) multimodales Modell (Anzahl Studien/ Vergleiche)	Anwender Ärzte/Schwestern/ andere Gesundheitsberufe	Präventionsstufe Primär/ Sekundär/ Tertiär	Design RCT/c-RCT	Untersuchung der Arztadhärenz als HZK/ NZK	Nachbeobachtungszeit 3-12 /> 12 Monate
Anwender-Erinnerungs- Systeme	1,30 (1,17 - 1,45) (n = 15)	**1,07 (0,93 - 1,23)** (n = 22/23)	22 (100%)/7 (32%)/5 (23%)	4 (18%)/18 (82%)/7 (32%)	3 (14%)/19 (86%)	17 (78%)/5 (22%)	20 (91%)/2 (9%)
Unterstützung von Datenflüssen	2,01 (1,02 - 3,96) (n = 1)	n.b.	1 (100%)/0 (0%)/0 (0%)	0 (0%)/1 (100%)/0 (0%)	0 (0%)/1 (100%)	0 (0%)/1 (100%)	1 (100%)/0 (0%)
Audit und Feedback	n.b.	**1,01 (0,73 - 1,40)** (n = 12/12)	12 (100%)/1 (8%)/3 (25%)	6 (50%)/9 (76%) /1 (8%)	2 (17%)/10 (83%)	7 (58%)/5 (42%)	9 (75%)/3 (25%)
Anwender-Fortbildungs-Maßnahmen	1,40 (1,16 - 1,69) (n = 15)	**1,34 (1,08 - 1,65)** (n = 25/29)	23 (92%)/4 (16%)/4 (14%)	5 (20%)/15 (60%)/7 (28%)	1 (4%)/24 (96%)	16 (64%)/9 (36%)	19 (76%)/6 (24%)
Patienten-Schulung	1,78 (1,00 - 3,16) (n = 4)	**1,48 (1,08 - 2,01)** (n = 14/15)	11 (79%)/6 (46%)/4 (31%)	3 (23%)/10 (77%)/2 (14%)	5 (36%)/9 (64%)	3 (21%)/11 (79%)	10 (71%)/4 (29%)
Unterstützung des Patienten-Selbst-Managements	1,16 (0,93 - 1,45) (n = 5)	**1,08 (0,80 - 1,45)** (n = 12/12)	10 (83%)/4 (33%)/1 (8%)	1 (8%)/9 (75%)/3 (25%)	7 (58%)/5 (42%)	1 (8%)/11 (92%)	12 (100%)/0 (0%)
Patienten-Erinnerungshilfen	n.b.	**0,81 (0,51 - 1,28)**	7 (100%)/4 (58%)/0 (0%)	0 (0%)/5 (83%)/2 (29%)	5 (71,4%)/2 (28,6%)	1 (14%)/6 (86%)	5 (71%)/2 (28%)
Organisatorische Veränderungen im Versorgungsablauf	1,96 (1,40 - 2,75) (n = 14)	**1,49 (1,21 - 1,82)**	15 (88%)/7 (41%)/12 (71%)	5 (29%)/14 (82%)/5 (29%)	8 (47%)/9 (53%)	6 (35%)/11 (65%)	13 (76%)/4 (24%)

n: Anzahl; RCT: individuell randomisierte Studie, c-RCT: cluster-randomisierte Studie, HZK: Hauptzielkriterium, NZK: Nebenzielkriterium, n.b.: nicht berechnet

Die hohe statistische Heterogenität in den Metaanalysen der unimodalen und multimodalen Strategien (τ^2=0,1899) erforderte Heterogenitätsanalysen, um die Effektmodifikation durch klinische und methodische Studieneigenschaften quantifizieren und die Genauigkeit der Schätzer verbessern zu können.

Dazu beschrieben wir zuerst den Zusammenhang zwischen der Häufigkeit negativer, geringer, moderater und hoher Behandlungseffekte und den Ausprägungen der Effektmodifizierer in einer Kreuztabelle (Table 2 in Unverzagt 2014c).

Die anschließenden Meta-Regressionsanalysen zeigten, dass im Vergleich zum Arzt als alleinigen Verantwortlichen der LL-Implementierung durch die Einbeziehung von Schwestern oder anderen Berufsgruppen die Arztadhärenz über alle eingesetzten Strategien verbessert werden konnte (ROR 1,29; 95 % KI 1,05-1,60 und ROR 1,62; 95 % KI 1,29-2,04). Die unerklärte Variabilität zwischen den Studien konnten wir durch die Einbeziehung dieses Effektmodifizierers um 27 % reduzieren (Tab. 4).

Tab. 4. Assoziation zwischen Effektmodifizierern und Behandlungseffekt über alle Kategorien der Implementierungsstrategien (aus Unverzagt 2014c, table 3).

Effektmodifizierer mit Vergleichsgruppen	**ROR (95%KI)**	**Zwischenstudien Variabilität τ^2; relative Reduktion der Varianz**
Zielgruppe Schwester vs. Arzt Andere Gesundheitsberufe vs. Arzt	1,29 (1,05 - 1,60) 1,62 (1,29 - 2,04)	0,1389; 26,9%
Primär vs. Tertiär-**Prävention** Sekundär-vs. Tertiär-Prävention	1,30 (0,98-1,71) 1,31 (1,09 - 1,57)	0,1692; 10,9%
Clusterrandomisierte RCT vs. RCT	1,28 (1,02 - 1,60)	0,1871; 1,5%
Anwender-Adhärenz als primärer vs. sekundärer **Endpunkt**	1,38 (1,12 - 1,70)	0,1719; 9,5%
Längere (>12 Monate) vs. kurze **Nachbeobachtungszeit**	1,38 (1,03 - 1,83)	0,1741; 8,3%
Verzerrungsrisiko (6 Bewertungen) hoch oder unklar vs. niedrig	n.b.	0,1488; 21,6%

ROR: Ratio of Odds Ratios; 95 % KI: 95 % Konfidenzintervall, n.b.: Es wurde kein gemeinsamer Einfluss über alle untersuchten Verzerrungsrisiken geschätzt.

Zusätzlich wiesen wir einen möglichen Einfluss der Präventionsebene nach, bei der Implementierungsstrategien in der Primär- und Sekundärprävention erfolgreicher als in der Tertiärprävention waren (ROR 1,30; 95 % KI 0,98-1,71 und ROR 1,31; 95 % KI 1,09-1,57).

Methodische Faktoren differenzierten zwischen Studien zur Verbesserung des Implementierungsprozesses (Prozessoptimierungsstudien) mit dem Hauptzielkriterium Arztadhärenz und Studien zur Optimierung von patientenrelevanten Endpunkten wie der

Mortalität, Morbidität oder weiterer, am Patienten erhobener Endpunkte (z.B. Blutdruck, Gesamtcholesterin, Body Mass Index). Wir konnten in Prozessoptimierungsstudien mit dem Hauptzielkriterium Arztadhärenz eine um 38 % höhere Arztadhärenz im Vergleich zu Studien mit patientenrelevanten Endpunkten identifizieren (ROR 1,38; 95 % KI 1,12-1,70). Ebenso zeigte sich eine positive Abhängigkeit der beobachteten Effekte von der Länge der Nachbeobachtungszeit und der Durchführung der Studie als cluster- und individual randomisierte kontrollierte Studie. Die einzelnen Verzerrungsquellen zeigten keinen signifikanten Einfluss auf den Behandlungseffekt, was wiederum durch die geringere Power beim Einbeziehen mehrerer Faktoren in die Regressionsanalysen bedingt sein könnte. Wir untersuchten die Effektmodifikation jedes untersuchten Faktors gemeinsam mit dem Einfluss von sieben unterschiedlichen Implementierungsstrategien auf den Behandlungseffekt und konnten deshalb keine Interaktionen zwischen den untersuchten Effektmodifizierern quantifizieren.

5. Diskussion

5.1 Methodische Erkenntnisse

Ziel meiner Arbeit war die Zusammenstellung methodischer Überlegungen, welche sich aus der Evidenzsynthese für konkrete klinische Fragestellungen ergaben. Diese betrafen den Umgang mit Verzerrungsquellen, nicht berichteten oder unpräzisen Behandlungseffekten, indirekter Evidenz aus Surrogat-Endpunkten und der Synthese von Behandlungseffekten aus Studien mit einer hohen Heterogenität. Diese Probleme ergaben sich aus Fragestellungen zur Behandlung des infarktbedingten kardiogenen Schocks (Unverzagt 2015) oder der Untersuchung von Implementierungsstrategien zur Verbesserung der ärztlichen Adhärenz zu Leitlinienempfehlungen (Unverzagt 2014b). Aus der Erfassung und Beschreibung von möglichen Verzerrungsquellen ergab sich die Fragestellung, in welchem Ausmaß diese zu systematisch verfälschten Behandlungseffekten führen können (Unverzagt 2013). Dazu existiert eine qualitative Beschreibung (Bellomo 2009), während eine quantitative meta-epidemiologische Untersuchung für das untersuchte Fachgebiet der Intensiv- und Notfallmedizin, für welches zur damaligen Zeit bereits weitere Übersichtsarbeiten (Unverzagt 2014e, Unverzagt 2015b) geplant waren, fehlten. Die Ergebnisse der hier vorgestellten meta-epidemiologischen Studie ermöglichen eine quantitative Abschätzung der Effektmodifikation durch Verzerrungsquellen in der Planung und Durchführung der Studien.

Eine weiteres methodisches Problem, welches ausführliche Heterogenitätsuntersuchungen erforderte, stellte sich in der Untersuchung des Einflusses von Implementierungsstrategien, deren Therapieeffekt durch verschiedene methodische und klinische Studieneigenschaften modifiziert wird (Unverzagt 2014c). Die resultierende Effektmodifikation soll hier diskutiert werden.

5.1.1 Untersuchung von Verzerrungsquellen

Auf der Grundlage von 82 Studien der Intensiv-und Notfallmedizin konnten wir aus insgesamt zwölf untersuchten Studieneigenschaften fünf Merkmale identifizieren, welche mit Überschätzungen und zwei, welche mit Unterschätzungen des Behandlungseffektes assoziiert waren (Unverzagt 2013). Zur Berechnung der Hauptanalysen nutzten wir ein REM-Modell, welches die durch die Diversität der untersuchten Interventionen, Definitionen der Ein- und Ausschlusskriterien, der Wahl der Kontrollgruppe

und der Länge der Nachbeobachtungszeiten entstehenden Unterschiede zwischen den beobachteten Behandlungseffekten berücksichtigte. Die Heterogenität zwischen den Metaanalysen reduzierten wir, einer Empfehlung von Welton (2009) folgend, durch die Begrenzung auf ein Fachgebiet, einen Endpunkt (Mortalität) und eine Datenbank.

Übereinstimmungen und Unterschiede zu anderen Studien

Wir beobachteten eine signifikante Überschätzung des Behandlungseffekt auf die Mortalität in monozentrischen im Vergleich zu multizentrisch durchgeführten Studien um 36 % (ROR 0,64; 95%KI 0,47-0,87). Damit bestätigt diese quantitative Untersuchung qualitative Beschreibungen von großen Behandlungseffekten in monozentrischen Studien der Intensiv- und Notfallmedizin, welche später in multizentrischen Studien widerlegt wurden (Bellomo 2009). Unsere Ergebnisse entsprechen den Ergebnissen meta-epidemiologischer Studien aus der Onkologie (Berlin 1989) und fachübergreifender Untersuchungen (Bafeta 2012, Dechartres 2011). Auch diese wiesen eine relative und absolute Überschätzung des Behandlungseffektes in monozentrischen Studien für binäre (ROR 0,73; 95%KI 0,64-0,83) und metrische Endpunkte (mittlere Differenz 0,09; 95%KI 0,01-0,17) nach. Im Gegensatz dazu konnte in einer weiteren fachübergreifenden Betrachtung von meta-epidemiologischen Studien kein signifikanter Unterschied zwischen den Behandlungseffekten mono- und multizentrisch durchgeführter Studien festgestellt werden, obwohl auch in dieser Arbeit monozentrische Studien zu höheren Behandlungseffekten tendierten (ROR 0,91; 95 % KI 0,79-1,04 in Savovic 2012).

Die Überschätzung des Behandlungseffektes in monozentrischen Studien korreliert mit dem sogenannten „small-study effect“. In kleinen Studien beobachtete Behandlungseffekte liegen häufig über denen in größeren Studien (Borenstein 2005, Sterling 1995, Sterne 2000). Gründe dafür werden darin gesehen, dass kleinere Studien häufig nur im Fall klinisch relevanter und signifikanter Behandlungseffekte von den Autoren veröffentlich werden (z.B. Easterbrook 1991, Sterling 1995, von Elm 2008), die Heterogenität der eingeschlossenen Patienten geringer ist (Sterne 2000) und häufig nur die Endpunkte in der Veröffentlichung beschrieben werden, welche signifikante Unterschiede aufweisen (Chan 2004a, 2004b, Williamson 2005). Auch wenn neue Online-Zeitschriften wie PLOS ONE, Scientific Reports und BMJ Open dem entgegenwirken, wenn sie methodisch einwandfrei bewertete Studien ohne Berücksichtigung der klini-

schen Relevanz ihrer Ergebnisse veröffentlichen, bleibt ein erhöhtes Risiko von Publikationsbias für kleine und monozentrisch durchgeführte Studien bestehen.

Wir empfehlen deshalb zusätzliche Subgruppenanalysen und die Berechnung spezifischer Effektschätzer für mono- und multizentrischen Studien in systematischen Übersichtsarbeiten und Evidenzzusammenstellungen; besonders wenn diese zur Generierung von LL-Empfehlungen angefertigt werden. Bei Effektunterschieden empfehlen wir eine stärkere Wichtung der Evidenz aus multizentrischen Studien mit hoher interner Validität, da diese die größere externe Validität aufweisen.

Die externe Validität ist bei monozentrischen Studien begrenzt. Diese werden häufig von hochmotivierten Medizinern mit ungewöhnlich hohen Fähigkeiten durchgeführt. Gerade RCTs in der Intensiv- und Notfallmedizin sind schwierig durchzuführen und teuer. Sie erfordern eine Auseinandersetzung mit organisatorischen, ethischen und juristischen Problemen, da die Patienten häufig selbst nicht in eine Studienteilnahme einwilligen können und viele Studien wegen Rekrutierungsschwierigkeiten vorzeitig beendet werden müssen (Thiele 2010). Außerdem unterscheiden sich Behandlungsmethoden in verschiedenen Ländern, die zur Verfügung stehenden Ressourcen, Fallzusammensetzung und die Kulturen im Umgang mit dem Lebensende (Bellomo 2009). Eine monozentrische Durchführung ist für patientenrelevante Endpunkte wie die Mortalität aufgrund der erforderlichen Fallzahl nicht realistisch, so dass Surrogat-Endpunkte zum Nachweis der Wirksamkeit gewählt werden. Alle diese Eigenschaften monozentrischer Studien können zu hohen Behandlungseffekten führen, welche sich nicht immer auf andere Situationen übertragen lassen (Windeler 2008) und damit die externe Validität monozentrischer Studien begrenzen.

Andererseits sehen wir außerhalb dieser Methodenkritik den hypothesengenerierenden Ansatz von monozentrischen RCTs als sehr hoch und wertvoll an, da anderenfalls viele Fragestellungen infolge des erhöhten Aufwands und der damit verbundenen Kosten nicht mehr oder nur noch dann aufgegriffen werden würden, wenn potente Geldgeber ein entsprechendes Finanzierungsinteresse sehen. Dieser Zusammenhang wird am Beispiel der in dieser Arbeit beschriebenen IABP-SHOCK-Studie deutlich. Ohne den hypothesengenerierenden Anstoß der kleinen monozentrischen RCTs wäre die nachfolgende Überprüfung eines seit mehr als 40 Jahren etablierten Therapiekonzeptes in einer multizentrischen RCT mutmaßlich nicht zustande gekommen.

Weitere Studieneigenschaften, welche möglicherweise mit einer Überschätzung des Behandlungseffektes assoziiert sind, sind das selektive Berichten von Endpunkten (ROR 0,80; 95 % KI 0,57-1,12), der Einsatz bestimmter Interventionen vor der Randomisierung (ROR 0,86; 95 % KI 0,73-1,00) und Strukturungleichheiten zum Zeitpunkt der Randomisierung (ROR 0,90; 95 % KI 0,75-1,07). Die von uns gezeigte Überschätzung von Studieneffekten bei einer möglicherweise selektiven Auswahl der berichteten Endpunkte bestätigt die Ergebnisse anderer meta-epidemiologischer Studien, dass klinisch relevante und statistisch signifikante Endpunkte häufiger publiziert werden (Chan 2004a, 2004b, Williamson 2005). In der Zusammenfassung der Behandlungseffekte in Metaanalysen kann daraus eine Überschätzung des Behandlungseffektes resultieren. Die Erfassung von Interventionen vor der Randomisierung, welche die folgende, randomisierte Intervention möglicherweise verstärken oder abschwächen können wird im Cochrane Handbuch empfohlen (The Cochrane Collaboration 2011; 8.15.1). So schlossen wir in unseren IABP-Review RCTs ein, in welchen Patienten teilweise bereits zum Zeitpunkt der Randomisierung mit einer IABP versorgt waren und danach in eine IABP oder eine Kontrollgruppe randomisiert wurden und vermuten eine Beeinflussung des resultierenden Behandlungseffektes. Aus diesem Grund werden in der Regel Patienten mit bestimmten, vor der Randomisierung erfolgten Interventionen, welche möglicherweise die Wirksamkeit einer zu untersuchenden Intervention beeinflussen, aus einer Studie ausgeschlossen. Diese Interventionen wurden von einem an der Übersichtsarbeit beteiligten Kardiologen vor Beginn der Bewertung der RCTs für jede untersuchte Intervention festgelegt. Auch die Untersuchung von Strukturungleichheiten zu Studienbeginn basiert auf einer Empfehlung des Cochrane Handbuchs (The Cochrane Collaboration 2011; 8.15.1) und von einem Kardiologen festgelegten, prognostisch relevanten Eigenschaften der Patienten. Viele der eingeschlossenen RCTs waren klein und wir betrachteten sehr viele Faktoren als prognostisch relevant, so dass wir in 45 % der eingeschlossenen Studien ungleiche Verteilungen dieser Faktoren identifizierten, welche den Behandlungseffekt bei fehlender Adjustierung beeinflussen können.

Im Allgemeinen wird von einer Überschätzung des Behandlungseffektes in Studien mit geringer methodischer Qualität ausgegangen (Egger 2003, Kjaergard 2001, Moher 1998, Savovic 2012, Schulz 1995, Wood 2008). Die existierende Evidenz ist aber nicht konsistent und variiert in Abhängigkeit von den untersuchten Interventionen, End-

punkten und der genauen Definition der untersuchten Studieneigenschaften. Es existieren Belege für eine Effektverzerrung des geschätzten Behandlungseffektes in beide Richtungen (z.B. Balk 2002, Odgaard-Jensen 2011). In der von uns durchgeführten Untersuchung identifizierten wir zwei Studieneigenschaften, welche zu einer Unterschätzung des Behandlungseffektes führen können. Eine unvollständige Beschreibung von Studienabbrüchen in der Nachbeobachtungszeit war mit einer mittleren Unterschätzung des Behandlungseffektes von 13 % (ROR 1,13; 95 % KI 0,89-1,42) assoziiert. Damit konnten wir die Ergebnisse einer meta-epidemiologischen Studie für onkologische Interventionen (Tierney 2005) nicht bestätigen, welche eine Effektüberschätzung zwischen 1 und 5 % beim Ausschluss von Patienten im Vergleich zu einer Auswertung auf Grundlage der vollständigen Intention-to-treat-population beschreiben.

Wir identifizierten einen möglicherweise verringerten Behandlungseffekt in Studien mit einer unvollständig oder unklar berichteten Beschreibung der verdeckten Therapiezuweisung (ROR 1,13; 95 % KI 0,94-1,19). Diese Beobachtung widerspricht den Ergebnissen einiger meta-epidemiologischer Arbeiten, welche eine Effektüberschätzung bei eingeschränkter Verdeckung der Therapiezuweisung beschreiben (Jüni 2001, Kjaergard 2001, Moher 1998, Schulz 1995,). Andere Studien konnten gerade für den in dieser Arbeit gewählten Endpunkt „Mortalität“ keinen Zusammenhang feststellen (Balk 2002, Savovic 2012, Wood 2008).

Keine Evidenz für eine Effektverzerrung konnten wir für die anderen sechs Verzerrungsquellen nachweisen, auch wenn diese in anderen meta-epidemiologischen Arbeiten als relevant erachtet wurden. Dies betraf beispielsweise den Einfluss vorzeitiger Studienabbrüche auf den Behandlungseffekt (ROR 1,10; 95 % KI 0,94-1,29). Wir konnten die Ergebnisse von Montori 2005 von extremeren Beobachtungseffekten in vorzeitig abgebrochenen Studien nicht bestätigen. Der Einfluss von Interessenkonflikten wird insgesamt widersprüchlich argumentiert. Einerseits werden industriegeförderte Studien teilweise nach hohen Qualitätsstandards durchgeführt (Kjaergard 2002) und andererseits können die Sponsoren die Interpretation und Veröffentlichung der Studienergebnisse beeinflussen (Als-Nielsen 2003, Bekelmann 2003, Gluud 2006, Gøtzsche 2005, Lexchin 2003), so dass sogar die Einführung eines „industry“ Bias empfohlen wird (Lundh 2012).

Andere Studieneigenschaften wie die Generierung der Randomisierung und die Methoden zur Verblindung werden häufig in aufeinanderfolgenden Sätzen im Kapitel ‚Methodik' der Studienberichte in vergleichbarer Qualität zur verdeckten Therapiezuweisung berichtet, so dass sich die Effekte dieser drei Eigenschaften schwer trennen lassen. Dies entspricht Beschreibungen von Wood et al. (2008), welche die unvollständige Darstellung der Methoden zur verdeckten Therapiezuweisung als Proxy für Studien mit einem hohem Verzerrungspotential bezeichnen. Ähnlich charakterisierten Schulz et al. (1996) die vollständige Darstellung von Studienabbrüchen in der Nachbeobachtungszeit, unsere zweite Studiencharakteristik, welche mit einer möglichen Unterschätzung des Behandlungseffektes assoziiert war. Auch die Häufigkeit von Therapiewechseln (cross-over) kann durch eine adäquate doppelte Verblindung beeinflusst werden, die einen systematischen Wechsel der Behandlungsarme verhindert. Die letzte von uns untersuchte Studieneigenschaft (ausreichende Nachbeobachtungszeit) über mindestens 28 Tage oder bis zur Entlassung aus dem Krankenhaus basierte auf hohen Mortalitätsraten nach der Entlassung von der Intensivstation bis zu einem Jahr nach der Randomisierung (Annane 2005, Moerer 2009). Wir empfehlen, diesen für den Patienten im höchsten Maß relevanten Zeitraum in die Bewertung des Behandlungseffektes einzubeziehen, konnten aber keine Unterschiede in den Behandlungseffekten zwischen Studien mit kurzer und längerer Nachbeobachtungszeit nachweisen.

Limitationen

Die Aussagen dieser Arbeit werden durch den hohen Anteil von fehlenden Informationen zur Bewertung der untersuchten Studieneigenschaften eingeschränkt. Die Qualität der Studienberichte, auf deren Grundlage wir die Studieneigenschaften bewerteten, muss nicht unbedingt die Qualität der Studiendurchführung widerspiegeln (Soares 2004), so dass wir in einigen Studien die interne Validität möglicherweise nicht richtig beurteilen konnten.

Zusätzlich kann von einem hohen Anteil von Interaktionen zwischen einigen untersuchten Studieneigenschaften ausgegangen werden, denn wir beobachteten, dass monozentrische Studien in nahezu allen untersuchten Studieneigenschaften ein höheres Verzerrungsrisiko als multizentrische Studien aufweisen und Schwächen in der Durchführung und Beschreibung der beobachteten Studieneigenschaften miteinander-

korrelieren. Aus diesem Grund untersuchten wir in Sensitivitätsanalysen alle untersuchten Studieneigenschaften zusätzlich zum multivariaten Modell auch univariat hinsichtlich ihrer Assoziation zum Behandlungseffekt. Die Möglichkeit von Confounding erachten wir als gering, da die univariaten Betrachtungen die Ergebnisse der simultanen multivariaten Betrachtungen aller untersuchten Eigenschaften bestätigen.

Insgesamt beschränkten wir das Modell aufgrund der eingeschränkten Anzahl eingeschlossener Studien und der geringen Anzahl von Metaanalysen je Übersichtsarbeit auf drei Hierarchieebenen (Metaanalysen, Studien und Studienteilnehmer). Wir gingen von einer ähnlichen Methodik in den untersuchten Übersichtsarbeiten aus, da diese alle nach der durch das Cochrane Handbuch beschriebenen Methodik durchgeführt wurden und verzichteten deshalb auf die Einbeziehung der Übersichtsarbeit als vierte Hierarchieebene. Die Analyse von Wechselwirkungen beschränkten wir auf die Wechselwirkungen zwischen Behandlungsgruppe und Metaanalyse.

Weitere Limitationen werden im Fehlen eines adäquaten hierarchischen statistischen Modells zum Umgang mit multiplen Studieneigenschaften und deren Interaktionen (Welton 2009), im hypothesengenerierendem Ansatz und in der geringen externen Validität dieser Arbeit gesehen, da diese auf ausgewählten (in der Cochrane Library publizierten) Metaanalysen basiert. Ich empfehle deshalb eine vorsichtige Interpretation der Ergebnisse dieser Untersuchungen.

5.1.2 Untersuchung von Heterogenität

In den vorgestellten Arbeiten untersuchten wir Ursachen der beobachteten statistischen Heterogenität der Beobachtungseffekte verschiedener RCTs in Subgruppen oder Metaregressionsanalysen, um so mögliche Effektmodifikationen durch klinische und methodische Studieneigenschaften erkennen und quantifizieren zu können. Hier sollen die Ergebnisse der Untersuchungen zur Wirksamkeit von Implementierungsstrategien auf die Adhärenz der behandelnden Allgemeinmediziner in der Behandlung von Patienten mit CVD (Unverzagt 2014c) diskutiert werden.

Übereinstimmungen und Unterschiede zu anderen Studien

Die Notwendigkeit ausführlicher Heterogenitätsuntersuchungen in dieser Übersichtsarbeit ergab sich aus der geringen Präzision und Konsistenz der beobachteten Behandlungseffekte. Die heterogenen Behandlungseffekte entsprechen den Ergebnissen aus

bereits existierenden fachübergreifenden Untersuchungen von Implementierungsstrategien (Grimshaw 2006, Grol 2003).

Klinische Faktoren, welche Behandlungseffekte modifizieren können, umfassen Eigenschaften der Studienpopulation, der untersuchten Intervention, der Vergleichsgruppen und erhobenen Endpunkte. Ihre Untersuchung kann helfen, den Prozess einer erfolgreichen LL-Implementierung genauer zu verstehen und Implementierungshilfen entsprechend anzupassen.

Effektmodifizierende Eigenschaften der Studienpopulation sahen wir in der Zielgruppe der Implementierungsstrategie und der Präventionsstufe der an CVD erkrankten Patienten. Ein großer Anteil der statistischen Heterogenität (27 %) konnte durch die Organisationsstrukturen in den allgemeinmedizinischen Praxen erklärt werden. Die LL-Konformität der Behandlung kann gesteigert werden, wenn Ärzte die Unterstützung durch andere Berufsgruppen wie Pharmazeuten (ROR 1,62; 95 % KI 1,29-2,04) oder qualifizierte Schwestern (ROR 1,29; 95 % KI 1,05-1,60) akzeptieren. Diese Ergebnisse bestätigen unsere Schlussfolgerungen aus der univariaten Untersuchung zur Wirksamkeit von Implementierungsstrategien (Unverzagt 2014b) und der meta-epidemiologischen fachübergreifenden Auswertung von Grol 2003, welche die Verantwortung des gesamten Praxisteams für die Implementierung von Leitlinien betonen.

Zusätzlich wurden Unterschiede in der LL-gerechten Versorgung zwischen den verschiedenen Präventionsstufen mit einer um ca. 30 % geringeren Adhärenz in der Tertiärprävention deutlich. Diese Unterscheidung konnte 11 % der unerklärten Variabilität der Behandlungseffekte erklären. Patienten leiden bei fortgeschrittenen Erkrankungen sowie zunehmendem Alter häufig unter komplexen Komorbiditäten, für deren Behandlung wenig Evidenz vorliegt, da diese Patienten häufig aus RCTs ausgeschlossen werden (Bailey 1994; Francke 2008). Dies bestätigt die Notwendigkeit von „effectiveness“ Studien, die untersuchen, inwieweit die Therapieeffekte aus RCTs auf Alltagsbedingungen in der Tertiärprävention von CVD übertragbar sind (Windeler 2008).

Die untersuchten Interventionen kategorisierten wir in Implementierungsstrategien, so dass Aussagen für die Wirksamkeit der verschiedenen Strategien möglich sind. Diese konnten die Arztadhärenz verbessern, wenn dafür organisatorische Veränderungen (ROR 1,49; 95 % KI 1,21-1,82) im Versorgungsablauf, Patienten- oder Anwenderschu-

lungen (ROR 1,48; 95 % KI 1,08-2,01 und ROR 1,34; 95 % KI 1,08-1,65) genutzt wurden. Ein längerer Einsatz der Strategien führte dabei zu einer verbesserten LL-Konformität (ROR 1,38; 95 % KI 1,04-1,83). Die Unterscheidung zwischen Studien mit kurzer und langer Nachbeobachtungszeit von über zwölf Monaten konnte 8 % der statistischen Heterogenität der Behandlungseffekte erklären. Längere Anwendungszeiten sind mit einer Wissenszunahme, der Möglichkeit einer verbesserten Integration der LL-Empfehlungen in die organisatorischen Strukturen der Praxen und der verbesserten Möglichkeit zur Überwindung von Vorbehalten verbunden.

Weitere, sowohl den Implementierungsprozess als auch die Verbesserung patientenrelevanter Endpunkte beeinflussende, hier aber nicht untersuchte klinische Faktoren, werden in Unterschieden zwischen den Gesundheitssystemen, der Art und Evidenzstärke der empfohlenen Maßnehmen und der teilweise dadurch begründeten Präferenzen des ärztlichen Teams und der Patienten gesehen (Ploeg 2007, Prior 2008). Alle diese Faktoren können sowohl die Präzision der Behandlungseffekte in den Einzelstudien als auch deren Konsistenz zwischen den Studien beeinflussen.

Unterschiede in der Studienmethodik der Einzelstudien sahen wir in der Wahl des Hauptzielkriteriums, der Länge der Nachbeobachtungszeit, der Durchführung mit einer individuellen oder Cluster-Randomisierung und im Auftreten potentieller Verzerrungsrisiken.

Beim kritischen Lesen der Volltexte potentiell geeigneter Studien stellten wir fest, dass verschiedene Typen von Hauptzielkriterien in den eingeschlossenen Studien festgelegt wurden. Eine Einbeziehung dieses methodischen Faktors konnte 10 % der statistischen Heterogenität der Behandlungseffekte erklären. Den festgelegten Einschlusskriterien genügten zum einen Studien, welche primär zur Verbesserung des Implementierungsprozesses mit dem Hauptzielkriterium „Arztadhärenz“ geplant wurden. Diese erfassten die Arztadhärenz über multiple Faktoren oder Scores, welche das Wissen, die Dokumentation, Aufklärung und Kontrolle von Risikofaktoren, die Verschreibung von Medikamenten und die Durchführung von diagnostischen Tests umfassten. Dabei kann Arztadhärenz selbstverständlich nur als Prozessparameter auf dem Weg zu einer Verbesserung patientenorientierter Endpunkten betrachtet werden. Andere eingeschlossene Studien untersuchten als primäres Ziel gerade die Verbesserung von patientenorientierten Endpunkten wie der Lebensqualität, Morbidität und Mortalität oder von Surro-

gat-Parametern wie dem Erreichen von Zielwerten oder der Reduzierung des kardiovaskulären Risikos und berichten Arztadhärenz als Prozessparameter. Prozessoptimierungsstudien erreichten im Vergleich zu Studien mit einem patientenrelevanten Hauptzielkriterium eine stärkere Erhöhung der Arztadhärenz (ROR 1,38; 95 % KI 1,12-1,70).

Voraussetzung für Effektivität von Leitlinienimplementierung im Hinblick auf patientenorientierte Endpunkte ist neben der verbesserten Arztadhärenz auch eine verbesserte Patientenadhärenz. Diese umfasst sowohl die Einnahmetreue von Medikamenten und als auch verhaltensmodifizierende Maßnahmen (Simpson 2006), deren Verbesserung wir in einer neuen Übersichtsarbeit für die Subgruppe der Patienten mit Herzinsuffizienz untersuchen (Unverzagt 2014a). Fehlende Patientenadhärenz bei Patienten mit CVD ist bei der Verschreibung von komplexen Therapieregimes häufig (Baroletti 2010) und verursacht einen großen Anteil der vermeidbaren stationären Aufnahmen (Laufs 2011).

Prozessoptimierungsstudien wurden in der Regel clusterrandomisiert durchgeführt und patientenrelevante Endpunkte werden in individuell randomisierten Studien erfasst, so dass es schwer ist, die Wahl des Hauptzielkriteriums von der gewählten Randomisierungseinheit getrennt zu interpretieren.

Die letzte untersuchte methodische Studieneigenschaft, welche Einschätzungen zur internen Validität der Studien umfasste, beeinflusste die Variabilität der beobachteten Effekte und konnte insgesamt 22 % der statistischen Heterogenität erklären, auch wenn wir keine einzelne potentielle Verzerrungsquelle mit einem signifikanten Zusammenhang zum Behandlungseffekt identifizieren konnten.

Limitationen

Methodische Überlegungen (Perleth 2012, Oxman 1992) betonen die Bedeutung, nur wenige, im Protokoll festgelegte Hypothesen zu testen. Diese frühen Festlegungen von Effektmodifikatoren basieren dann aber häufig auf pathophysiologischen Überlegungen oder der Kenntnis ausgewählter Studien mit hohen Behandlungseffekten in diesen Subgruppen. In vielen SRs werden im Protokoll viele Subgruppen festgelegt, für welche dann in den Studien keine spezifischen Ergebnisse berichtet oder Individualdaten zur Verfügung gestellt werden, so dass deren Untersuchung in Subgruppen- oder Metare-

gressionsanalysen nicht möglich ist. Wir ergänzten unsere frühen Festlegungen von Subgruppen im Protokoll nach dem Lesen aller Volltexte, aber vor der Datenextraktion, auch wenn die darauf basierenden Analysen dann Ergebnis einer frühen Datenexploration sind. In der hier diskutierten Arbeit basiert die Evidenz der Unterschiede zwischen den Subgruppen neben der Höhe der beobachteten Unterschiede und dem Anteil der erklärten Varianz immer auf plausiblen methodischen und soziologischen Erklärungen.

Kritisch bewerten wir auch die zusammenfassende Erfassung und Auswertung von LL-Konformität über mehrere Handlungsfelder wie Beratung, Diagnostik und Therapieempfehlungen. Wir konnten die Felder nicht trennen, da in den Originalarbeiten häufig zusammenfassende Scores berichtet wurden, welche Arztadhärenz in verschiedenen Handlungsfeldern zusammenfassen. Aus diesem Grund geben wir keine spezifischen Empfehlungen für die einzelnen Handelsfelder, an welchen LL-Konformität gemessen wurde.

Auch die geplante Untersuchung des Zusammenhangs zwischen dem von uns vollständig erfasstem Parameter Arztadhärenz und patientenrelevanten Endpunkten erwies sich als unmöglich, da diese aufgrund der notwendigen langen Nachbeobachtungszeiten gerade in der Primär- und Sekundärprävention von CVD nur unregelmäßig erfasst und berichtet wurden.

Die Ergebnisse der hier diskutierten Heterogenitätsanalysen bieten die Möglichkeit, den Prozess einer erfolgreichen LL-Implementierung genauer zu verstehen, sollten aber auf jeden Fall hypothesengenerierend interpretiert werden, um falsch-positive Ergebnisse zu vermeiden (Higgins 2004, Oxman 1992).

5.2 Auswirkungen auf die klinische Praxis in der Behandlung des kardiogenen Schocks

Der Einsatz der IABP war bei Patienten mit infarktbedingtem kardiogenen Schock gemeinsam mit einer PTCA oder einer Bypassoperation eine Klasse 1-Empfehlung der American Heart Association (AHA) und der European Society of Cardiology (ESC) - Leitlinien (Antman 2004, van der Werf 2008). Diese Empfehlung ist in den Leitlinien für die Behandlungsoptionen vorgesehen, für welche ausreichend Evidenz oder eine allgemeine Übereinstimmung vorliegen, dass ihr Einsatz nützlich, sinnvoll und wirksam ist und der erwartete Nutzen deutlich den Schaden überwiegt. Die Empfehlung der AHA

basiert auf einem Evidenzlevel B mit einer Evidenz aus einer RCT für eine ähnliche Indikation und aus nichtrandomisierten Studien bei Patienten mit kardiogenem Schock (Antman 2004). Die ESC vergab den Evidenzlevel C auf Grundlage eines Expertenkonsens und von Daten aus kleinen oder retrospektiv durchgeführten Studien (van der Werf 2008).

Im Ergebnis unserer und einer weiteren systematischen Übersichtsarbeit, welche neben RCTs auch nichtrandomisierte, prospektiv geplante Studien mit insgesamt 10529 Schockpatienten einschließt (Sjauw 2009), wurden die Empfehlungen in den amerikanischen, europäischen und deutschen Leitlinien zum Einsatz der IABP in der Behandlung des infarktbedingten kardiogenen Schocks geändert. Der Routineeinsatz einer IABP wird jetzt auf Grundlage einer hohen Evidenz (Level A) beim infarktbedingtem kardiogenen Schock nicht länger empfohlen (O'Gara 2013, Steg 2012). Auch die starke Empfehlung in den deutsch-österreichischen Leitlinien wurde herabgestuft (Werdan 2012) und zurzeit unter Nutzung der Ergebnisse der multizentrischen Studie von Thiele (2012) und der Ergebnisse unserer Übersichtsarbeit überarbeitet (persönliche Mitteilung von Prof. Werdan, März 2015). Infolge dieser Empfehlungsänderungen sank die Häufigkeit des IABP-Einsatzes in Deutschland und Österreich im Zeitraum von 2008 bis 2012 um 17 % (Healthcare in Europe 2014). Dieser Rückgang wird für den nachfolgenden Betrachtungszeitraum ab dem Jahre 2013 als noch wesentlich ausgeprägter erwartet.

5.3 Schlussfolgerungen

Aus dem IABP-Review kann auf der Grundlage einer multizentrisch und mehrerer monozentrisch durchgeführter Studien keine Evidenz für einen routinemäßigen Einsatz einer IABP bei Patienten mit infarktbedingtem kardiogenen Schock abgeleitet werden.

Eine zusammenfassende Untersuchung mehrerer Metaanalysen aus dem Fachgebiet der Intensiv- und Notfallmedizin lieferte einen Beleg für den Einfluss der mono- oder multizentrischen Durchführung einer Studie auf die beobachtete Effektstärke mit höheren Effekten in monozentrisch durchgeführten Studien. Der Einfluss aller anderen Studieneigenschaften auf den Behandlungseffekt kann nur hypothesengenerierend interpretiert werden.

Die Ergebnisse des SRs zur Wirksamkeit von Implementierungsstrategien zeigen, dass sowohl uni- als auch multimodale Strategien zu einer Verbesserung der Arztadhärenz führen können. Strategien, welche organisatorische Veränderungen im Versorgungsablauf, nicht-ärztliche Gesundheitsberufe sowie Patienten- und Anwenderschulungen einbeziehen, scheinen besonders erfolgreich zu sein. Das gesamte medizinische Personal in Praxen und Kliniken sollte in die Umsetzung der Implementierungsstrategien einbezogen werden. Eine verbesserte Arztadhärenz wird als wichtiger Prozessparameter gesehen, der aber nicht unbedingt auch zu einer verbesserten Patientenadhärenz und besseren patientenrelevanten Endpunkten führen muss. Der Effekt der unterschiedlichen Implementierungsstrategien kann durch klinische Faktoren wie den Adressaten der Strategie, die Patientenpopulation und den Zeitraum, über welchen hinweg diese eingesetzt wird, beeinflusst werden. Aber auch methodische Faktoren wie das Studiendesign und potentielle Verzerrungsquellen können den Behandlungseffekt einer Implementierungsstrategie modifizieren.

6. Zusammenfassung

SRs werden in verschiedenen Evidenzhierarchiesierungen mit der höchsten Evidenzstufe beurteilt. Aber häufig rechtfertigt die zusammengefasste Evidenz aus diesen Arbeiten keine starke LL-Empfehlung für oder gegen die untersuchten Interventionen.

Diese Arbeit stellt klinische Ergebnisse und methodische Überlegungen aus der Durchführung einer RCT, zweier SRs, einer meta-epidemiologischen Studie und ausführlicher Heterogenitätsbetrachtungen zusammenfassend dar und diskutiert diese. Sowohl die Studie als auch beide Übersichtsarbeiten wurden zum Nachweis der therapeutischen Überlegenheit der untersuchten therapeutischen Maßnahmen geplant.

Wir konnten keinen Nutzen für eine medizinisch etablierte Intervention (Einsatz einer IABP bei Patienten mit infarktbedingtem kardiogenen Schock) auf der Grundlage individueller Patientendaten nachweisen. Aus der Beschreibung und Beurteilung von Verzerrungsquellen in Studien der Intensiv- und Notfallmedizin ergab sich die Frage, in welchem Ausmaß Verzerrungsquellen in der Planung und Durchführung von RCTs und die mono- oder multizentrische Durchführung einer Studie einen Einfluss auf den Behandlungseffekt des SRs aufweisen. Eine systematische Verzerrung des Behandlungseffektes auf den Endpunkt Mortalität konnten wir für keine der untersuchten Verzerrungsquellen zeigen. Im Gegensatz dazu konnten wir eine Überschätzung des Behandlungseffektes um ca. 36 % (95 % KI 13-53 %) in monozentrischen im Vergleich zu multizentrisch durchgeführten Studien nachweisen.

Implementierungsstrategien konnten die LL-Konformität der Ärzte in der allgemeinmedizinischen Versorgung vor allem dann verbessern, wenn sie organisatorische Veränderungen im Versorgungsablauf und Maßnahmen zur Patienten- und Anwenderschulung einsetzten. Eine Einbindung nicht-ärztlicher Gesundheitsberufe könnte geeignet sein, vorhandene Barrieren in allgemeinmedizinischen Praxen gegenüber einer LL-gerechten Veränderung von Versorgungsabläufen zu reduzieren. Zusätzlich wurden Unterschiede in der LL-gerechten Versorgung zwischen den verschiedenen Versorgungsstufen mit einer geringeren Adhärenz in der Tertiärprävention deutlich. Methodisch wurde ein Einfluss der Planung der Studie mit höheren Effekten in Studien mit dem Hauptzielkriterium LL-Konformität nachgewiesen.

7. Referenzen

Als-Nielsen B, Chen W, Gluud C, Kjaergaard LL (2003) Association of funding and conclusions in randomized drug trials: a reflexion of treatment effects or adverse events? JAMA 290: 921-8.

Annane D, Bellissant E, Cavaillon JM (2005) Septic Shock. Lancet 365: 3-78.

Altman DG (1985) Comparability of randomised groups. Statistician 34:125-36.

Altman DG. Systematic reviews of evaluations of prognostic variables (2001) In: Egger M, Davey Smith G, Altman DG (Ed). Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing Group: 228-47.

Antes G, Blümle A, Lang B (2014) Medizinisches Wissen – Entstehung, Aufbereitung, Nutzung. In: Rebscher H, Kaufmann S (Hrsg.) Wissensmanagement im Gesundheitswesen, Heidelberg: Medizinrecht.de Verlag: 29-49.

Antman EM, Anbe DT, Armstrong PW, Bates ER, Green LA, et al. (2004) ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction: executive summary: a report of the ACC/AHA Task Force on Practice Guidelines. J Am Coll Cardiol 44: 671–719.

AZQ (Ärztliches Zentrum für Qualität in der Medizin) (2007) Das LL--Glossar. Begrifflichkeiten und Kommentare zum Programm für Nationale Versorgungsleitlinien. Berlin: ÄZQ, 160 S. http://www.awmf.org/fileadmin/user_upload/Leitlinien/ Werkzeuge/ ll-glossar.pdf (Zugriff am 27.3.2015).

Babaev A, Frederick PD, Pasta DJ, Every N, Sichrovsky T, Hochman JS (2005) Trends in management and outcomes of patients with acute myocardial infarction complicated by cardiogenic shock. JAMA 294: 448–54.

Bafeta A, Dechartres A, Trinquart L, Yavchitz A, Boutron I, Ravaud P (2012) Impact of single center status on estimates of intervention effects in trials with continuous outcomes: meta-epidemiological study. BMJ 344: e813.

Bailey KR (1994) Generalizing the results of randomized clinical trials. Control Clin Trials 15: 15–23.

Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, et al. (2002) Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA 287: 2973-82.

Barron HV, Every NR, Parsons LS, Angeja B, Goldberg RJ, et al. (2001) The use of intra-aortic balloon counterpulsation in patients with cardiogenic shock complicating acute myocardial infarction: data from the National Registry of Myocardial Infarction 2. Am Heart J 141: 933–9.

Bellomo R, Warrillow SJ, Reade MC (2009) Why we should be wary of single-center trials. Crit Care Med 37: 3114-9.

Berlin JA, Begg CB, Louis TA (1989) An assessment of publication bias using a sample of published clinical trials. J Am Stat Assoc 84: 381-92.

Borenstein M, Hedges LV, Higgins JP, Rothstein HR (2009) Introduction to Meta-Analysis. Chichester: John Wiley & Sons, 421 S.

Burzykowski T, Buyse M, Piccart-Gebhart MJ, Sledge G, Carmichael J, et al. (2008) Evaluation of Tumor Response, Disease Control, Progression-Free Survival, and Time to Progression As Potential Surrogate End Points in Metastatic Breast Cancer. J Clin Oncol 26: 1987-92.

Chalmers I (2001) Comparing like with like: some historical milestones in the evolution of methods to create unbiased comparison groups in therapeutic experiments. Int J Epidemiol 30: 1156-64.

Campbell MK, Fayers PM, Grimshaw JM (2005) Determinants of the intracluster correlation coefficient in cluster randomized trials: the case of implementation research. Clin Trials 2: 99–107.

Campbell MK, Piaggio G, Elbourne DR, Altman DG, Group C (2012) Consort 2010 statement: extension to cluster randomised trials. BMJ 345: e5661.

Chan A-W, Hróbartsson A, Haahr MT, Gøtzsche PC, Altman DG (2004a) Empirical evidence for selective reporting of outcomes in randomized trials. Comparison of protocols to published articles. JAMA 291: 2457-65.

Chan AW, Krieca-Jeric K, Schmid I, Altman DG (2004b) Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. Can Med Assoc J 171: 735-40.

Cochrane Collaboration (2011) Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. Higgins JPT, Green S (Hrsg.) Oxford: The Cochrane Collaboration.

Dechartres A, Boutron I, Triquart L, Charles P, Ravaud P (2011) Single-center trials show larger treatment effect than multicenter trials: evidence from a meta-epidemiological study. Ann Intern Med 155: 39-51.

Deeks JJ (2002) Issues in the selection of a summary statistics for meta-analysis of clinical trials with binary outcomes. Stat Med 21: 1575-600.

Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR (1991) Publication bias in clinical research. Lancet 337: 867-72.

Egger M, Juni P, Bartlett C, Holenstein F, Sterne J (2003) How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 7: 1-76.

Egger M, Smith GD, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. BMJ 315: 629-34.

Fergusson D, Aaron SD, Guyatt G, Hebert P (2002) Post-randomisation exclusions:the intention to treat principle and excluding patients from analysis. BMJ 325: 652-4.

Francke AL, Smit MC, de Veer AJ, Mistiaen P (2008) Factors influencing the implementation of clinical guidelines for health care professionals: a systematic meta-review. BMC Med Inform Decis Mak 8: 38.

Giraudeau B, Ravaud P (2009) Preventing bias in cluster randomised trials. PLoS Med 6: e1000065.

Glasziou PP, Sanders SL (2002) Investigating causes of heterogeneity in systematic reviews. Stat Med 21: 1503–11.

Gluud LL (2006) Bias in clinical intervention research. Am J Epidemiol 163: 493-501.

Goodman S, Dickersin K (2011) Metabias: a challenge for comparative effectiveness research [Editorial]. Ann Intern Med 155: 61-2.

Gøtzsche PC (2005) Research integrity and pharmaceutical industry sponsorship [Editorial]. Med J Aust 182: 549-50.

Grimshaw J, Eccles M, Thomas R, Maclennan G, Ramsey G, et al. (2006) Toward evidence-based quality improvement. Evidence (and its limitations) of te evectiveness of guideline dissemination and implementation strategies 1966-1998. J Gen Intern Med 21: S14-20.

Groll R, Grimshaw J (2003) From best evidence to best practice: effective implementation of change in patient care. Lancet 362: 1225-30.

Guyatt G, Oxman AD, Akl E, Kunz R, Vist G, et al. (2011a) GRADE guidelines 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 64: 383-94.

Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, et al. (2011b) GRADE guidelines 4: rating the quality of evidence - risk of bias. J Clin Epidemiol 64: 407-15.

Guyatt G, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, et al. (2011c) GRADE guidelines 6. Rating the quality of evidence - imprecision. J Clin Epidemiol 64: 1283-93.

Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, et al. (2011d) GRADE guidelines: 7. Rating the quality of evidence - inconsistency. J Clin Epidemiol 64: 1294-1302.

Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, et al. The GRADE Working Group. (2011ce) GRADE guidelines: 8. Rating the quality of evidence - indirectness. J Clin Epidemiol 64: 1303-10.

Hasenbein U, Wallesch C-W (2007) Was ist Leitlinienkonformität? Theoretische und methodische Überlegungen zu einem aktuellen Konzept der Versorgungsforschung und Qualitätssicherung. Gesundheitswesen 69: 427-37.

Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, et al. (2009) Risk of bias versus quality assessment of randomized controlled trials: cross sectional study. BMJ 339: b4012.

Healthcare in Europe (2014) Shocked by the shock trials? http://www.healthcare-in-europe.com/en/article/12105-Shocked_by_the_Shock_trials.html. (Zugriff am 16.4.2015).

Higgins JP, Thompson SG (2002) Quantifying heterogeneity in a metaanalysis. Stat Med 21: 1539–58.

Higgins JP, Thompson SG (2004) Controlling the risk of spurious findings from meta-regression. Stat Med 23: 1663-82.

Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327: 557–60.

Higgins J, Thompson S, Deeks J, Altman D (2002) Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. J Health Serv Res Policy 7: 51–61.

ICH E9 (1998) Statistical principles for clinical trials. ICH Expert Working Group. http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/statistical-principles-for-clinical-trials.html (Zugriff am 21.3.2015).

Jüni P, Altman DG, Egger M (2001) Assessing the quality of controlled clinic trials. BMJ 323: 42-6.

Kjaergard LL, Villumsen J, Gluud C (2001) Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 135: 982-9.

Klement A, Oemler M, Braun K, Unverzagt S (2015) Strategien zur Implementierung von Leitlinien. Z Allg Med 91(2).

Kuss O (2002) How to use SAS for logistic regression with correlated data. Proc 27th Annu SAS Users Group Int Conf (SUGI 27): 261-27.

Laufs U, Böhm M, Kroemer HK, et al. (2011) Strategien zur Verbesserung der Einnahmetreue von Medikamenten. Dtsch Med Wochenschr 136: 1616–21.

Lefante J, Krousel-Wood M (2006) Overcoming barriers: the role of providers in improving patient adherence to antihypertensive medications. Curr Opin Cardiol 21: 310-315.

Lundh A, Sismondo S, Lexchin J, Busuioc OA, Bero (2012) Industry sponsorship and research outcome. Cochrane Database Syst Rev 12.

Moerer O, Quintel M (2009) Epidemiologie und ökonomische Aspekte der Sepsis bei Erwachsenen. Internist 50: 788-98.

Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, et al. (2001) CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 63: e1-e37.

Moher D, Liberati A, Tetzlaff J, Altman DG, the PRISMA Group (2009) Preferred Reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLOS Mrd 6: e1000097.

Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. (1998) Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 352: 609-13.

Moher D, Schulz KF, Altman DG, for the CONSORT Group (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 357: 1191–4.

Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, et al. (2005) Randomized trials stopped early for benefit: a systematic review. JAMA 294: 2203-9.

OCEBM Levels of Evidence Working Group*. "The Oxford 2011 Levels of Evidence". Oxford Centre for Evidence-Based Medicine. http://www.cebm.net/index.aspx?o=5653 (Zugriff am 28.3.2015).

Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, et al.(2011) Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev 4.

O'Gara PT, Kushner FG, Ascheim DD, Casey DE Jr, Chung MK, et al. (2013) 2013 ACCF/AHA Guideline for the Management of ST-Elevation Myocardial Infarction: A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation 127: 529-55.

Ollenschläger G, Meyer G, Windeler J (2014) 20 Jahre Leitlinien und EbM in Deutschland: Ist EbM in der Versorgung angekommen? Z Evid Fortbild Qual Gesundhwes 10: 547-8.

Oxman AD, Guyatt GH (1992) A Consumer's guide to subgroup analyses. Ann Intern Med 116: 78-84.

Parmar MKB, Torri V, Steward L (1998) Extracting summary statistics to perform meta-analysis of the published literature for survival endpoints. Stat Med 17: 2814–34.

Perleth M, Langer G, Meerpohl JJ, Gartlehner G, Kaminsky-Hartenthaler A, Schünemann HJ (2012) GRADE Leitlinien: 7. Einschätzung der Qualität der Evidenz – Inkonsistenz. ZEFQ 106: 733-44.

Piedbois P, Mondor H, Buyse M (2004) Meta-analyses based on abstracted data: A step in the right direction, but only a first step. J Clin Oncol 22: 3839–41.

Pigott T, Shepperd S (2013) Identifying, documenting, and examining heterogeneity in systematic reviews of complex interventions. J Clin Epidemiol 66: 1244–50.

Ploeg J, Davies B, Edwards N, Gifford W, Miller PE (2007) Factors influencing best-practice guideline implementation: lessons learned from administrators, nursing staff, and project leaders. Worldviews Evid Based Nurs 4: 210–219.

Prior M, Guerin M, Grimmer-Somers K (2008) The effectiveness of clinical guideline implementation strategies–a synthesis of systematic review findings. J Eval Clin Pract 14: 888–897.

Prondzinsky R, Lemm H, Swyter M, Wegener N, Unverzagt S, et al. (2010) Intra-aortic balloon counterpulsation in patients with acute myocardial infarction complicated by cardiogenic shock: the prospective, randomized IABP SHOCK Trial for attenuation of multiorgan dysfunction syndrome. Crit Care Med 38: 152-60.

Prondzinsky R, Unverzagt S, Russ M, Lemm H, Swyter M, et al. (2012a) Hemodynamic Effects of Intra-aortic Balloon Counterpulsation in Patients With Acute Myocardial Infarction Complicated by Cardiogenic Shock: The Prospective, Randomized IABP Shock Trial. Shock 37: 378-84.

Prondzinsky R, Unverzagt S, Lemm H, Wegener NA, Schlitt A, et al. (2012b) Interleukin-6, -7, -8 and -10 predict outcome in acute myocardial infarction complicated by cardiogenic shock. Clin Res Cardiol 101: 375-84.

Prondzinsky R, Unverzagt S, Lemm H, Wegener N, Heinroth K, et al. (2012c) Acute myocardial infarction and cardiogenic shock: Prognostic impact of cytokines: INF-γ, TNF-α, MIP-1β, G-CSF, and MCP-1β. Med Klin Intensivmed Notfmed 107: 476–84.

Puffer S, Torgerson D, Watson J (2003) Evidence for risk of bias in clusterrandomised trials: review of recent trials published in three general medical journals. BMJ 327: 785–9.

Review Manager (RevMan) [Computer program]. Version 5.3. (2014) Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration.

Riley RD, Simmonds MC, Look MP (2007) Evidence combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. J Clin Epidemiol 60: 431–9.

Riley RD, Lambert PC, Abo-Zaid G (2010) Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ 340: 521–5.

Rosén M (2009) The aprotinin saga and the risks of conducting meta-analyses on small randomised controlled trials - a critique of a Cochrane review. BMC Health Serv Res 9: 34.

Rucker G, Schwarzer G, Carpenter JR, Schumacher M (2008) Undue reliance on I^2 in assessing heterogeneity may mislead. BMC Med Res Methodol 8: 79.

Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS (1996) Evidence based medicine: what it is and what it isn't. BMJ 312: 71-2

Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, et al. (2012) Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med 157: 429-38.

Schulz KF, Altman DG, Moher D, Group C (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. BMJ 340: c332.

Schulz KF, Chalmers I, Hayes RJ, Altman DG (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 273: 408–12.

Schulz KF, Grimes DA, Altman DG, Hayes RJ (1996) Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. BMJ 312: 742-4.

Shojania KG, McDonald KM, Wachter RM, Owens DK (2004). Closing The Quality Gap: A Critical Analysis of Quality Improvement Strategies, Volume 1—Series Overview and Methodology. Technical Review 9 (Contract No. 290-02-0017 to the Stanford University–UCSF Evidence-based Practices Center). AHRQ Publication No. 04-0051-1. Rockville, MD: Agency for Healthcare Research and Quality, 48 S.

Siersma V, Als-Nielsen B, Chen W, Hilden J, Gluud LL, et al (2007) Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. Stat Med 26: 2745-58.

SIGN (Scottish Intercollegiate Guidelines Network) (2008) SIGN 50. A guideline developer's handbook. Revised edition. Edinburgh, www.sign.ac.uk, 112 S.

Simpson SH, Eurich D, Majumdar SR, Padwal RS, Tsuyuki RT, et al. (2006) A meta-analysis of the association between adherence to drug therapy and mortality. BMJ 333: 15.

Smith CT, Williamson PR, Marson AG (2005) Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. Stat Med 24: 1307–19.

Soares HP, Daniels S, Kumar A, Clarke M, Scott C, et al. (2004) Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. BMJ 328: 22-4.

Sterne JA, Gavaghan D, Egger M (2000) Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol 53: 1119-29.

Steg G, James SK, Atar D, Badano LP, Blomstrom Lundqvist C, et al. (2012) ESC GUIDELINES: ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force on the management of ST-

segment elevation acute myocardial infarction of the European Society of Cardiology (ESC). Eur Heart J 33: 2569-619.

Sterling TD, Rosenbaum WL, Weinkam J (1995) Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. Amer Statist 49: 108-12.

Sterne JA, Gavaghan D, Egger M (2000) Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol 53: 1119-29.

Stewart LA, Clarke MJ on behalf of the Cochrane working group on meta-analysis using individual patient data (1985). Practical methodology of meta-analyses (overviews) using updated individual patient data. Stat Med 14, 2057-79.

Thiele H, Allam B, Chatellier G, Schuler G, Lafont A (2010) Shock in acute myocardial infarction: the cape horn for trials? Eur Heart J 31: 1828-35.

Thiele H, Zeymer U, Neumann F-J, Ferenc M, Olbricht H-G, et al. (2012) Intraaortic balloon support for myocardial infarction with cardiogenic shock. NEJM 367: 1287-96.

Thom T, Haase N, Rosamund W, Howard VJ, Rumsfeld J, et al. (2006) Heart disease and stroke statistics – 2006 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 113: e85-151.

Thompson SG (1994) Why sources of heterogeneity in meta-analysis should be investigated. BMJ 309: 1351–5.

Tierney JF, Stewart LA (2005) Investigating patient exclusion bias in metaanalysis. Int J Epidemiol 34: 79-87.

Tierney JF, Stewart L, Ghersi D, Burdett S, Sydes MR (2007) Practical methods for incorporating summary time-to-event data into meta-analysis. Trials 8: 16.

Unverzagt S, Buerke M, de Waha A, Haerting J, Pietzner D, et al. (2015) Intra-aortic balloon pump counterpulsation (IABP) for myocardial infarction complicated by cardiogenic shock. Cochrane Database Syst Rev 3.

Unverzagt S, Hirsch K, Prondzinsky R (2015b) Vasopressors and predominantly vasoconstrictive drugs for acute myocardial infarction complicated by cardiogenic shock [Protocol]. Cochrane Database Syst Rev 3.

Unverzagt S, Klement A, Meyer G, Prondzinsky R (2014a) Interventions to enhance adherence to guideline recommendations in secondary and tertiary prevention of heart failure: A systematic review. J Clin Trials 4: 6.

Unverzagt S, Machemer M-T, Solms A, Thiele H, Burkhoff D, et al. (2011) Intra-aortic balloon pump counterpulsation (IABP) for myocardial infarction complicated by cardiogenic shock. Cochrane Database Syst Rev 7.

Unverzagt S, Oemler M, Braun K, Klement A (2014b) Strategies for guideline implementation in primary care focusing on patients with cardiovascular disease: a systematic review. Fam Pract 31: 247–66.

Unverzagt S, Peinemann F, Oemler M, Braun K, Klement A (2014c) Meta-regression analyses to explain statistical heterogeneity in a systematic review of strategies for guideline implementation in primary care. PLoS ONE 9: e110619.

Unverzagt S, Prondzinsky R, Peinemann F (2013) Single-center trials tend to provide larger treatment effects than multicenter trials: a systematic review. J Clin Epidemiol 66: 1271-80.

Unverzagt S, Wachsmuth L, Hirsch K, Thiele H, Buerke M, et al. (2014d) Inotropic agents and vasodilator strategies for acute myocardial infarction complicated by cardiogenic shock or low cardiac output syndrome. Cochrane Database Syst Rev 1.

Van de Werf F, Bax J, Betriu A, Blomstrom-Lundqvist C, Crea F, et al. (2008) Management of acute myocardial infarction in patients presenting with persistent ST-segment elevation: the Task Force on the Management of ST Segment Elevation Acute Myocardial Infarction of the European Society of Cardiology. Europ Heart J 29: 2909–45.

von Elm E, Rollin A, Blumle A, Huwiler K, Witschi M, Egger M (2008) Publication and non-publication of clinical trials: longitudinal study of applications submitted to a research ethics committee. Swiss Med Wkly 138: 197-203.

van Nieuwenhoven CA, Buskens E, van Tiel FH, Bonten MJ (2001) Relationship between methodological trial quality and the effects of selective digestive decontamination on pneumonia and mortality in critically ill patients. JAMA 286: 335-40.

Verhagen AP, de Vet HC, de Bie RA, Boers M, van den Brandt PA (2001) The art of quality assessment of RCTs included in systematic reviews. J Clin Epidemiol 54: 651-4.

Wagner AD, Thomssen C, Haerting J, Unverzagt S (2012) Vascular-endothelial-growth-factor (VEGF) targeting therapies for endocrine refractory or resistant metastatic breast cancer. Cochrane Database Syst Rev 7.

Wagner AD, Arnold D, Grothey A, Haerting J, Unverzagt S (2009) Anti-angiogenic therapy for metastatic colorectal cancer. Cochrane Database Syst Rev 3.

Welton NJ, Ades AE, Carlin JB, Altman DG, Sterne JAC (2009) Models for potentially biased evidence in meta-analysis using empirically based priors. J R Stat Soc Ser A 172: 119-36.

Werdan K, Pilz G, Bujdoso O, Fraunberger P, Neeser G, et al. (2007) Score-Based Immunoglobulin Therapy of Sepsis (SBITS) Study Group. Score-based immunoglobulin therapy of patients with sepsis: the SBITS study. Crit Care Med 35: 2693-2701.

Werdan K, Ruß M, Buerke M, Delle-Karth G, Geppert A, et al. (2012) Cardiogenic shock due to myocardial infarction: diagnosis, monitoring and treatment: a German-Austrian S3 Guideline. Dtsch Arztebl Intern 109: 343-51.

Whitehead, A. (2002) Meta-Analysis of Controlled Clinical Trials. John Wiley & Sons Ltd., England, 336 S.

WHO (2003) Adherence to long-term therapies: evidence for action. http://www.who.int/chp/knowledge/publications/adherence_report/ en/) (Zugriff am 16.3.2015).

Williamson BR, Gamble C (2005) Identification and impact of outcome selection bias in meta-analysis. Stat Med 24: 1547-61.

Windeler J (2008) Externe Validität. ZEFQ 102: 253-60.

Wood L, Egger M, Gluud LL, Schulz KF, Juni P, et al. (2008) Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: metaepidemiological study. BMJ 336: 601-5.

8. Thesen

1. Systematische Übersichtsarbeiten werden in Evidenzhierarchisierungssystemen mit der höchsten Evidenzstufe beurteilt. In dieser Arbeit zeigen wir an konkreten klinischen Beispielen, wann die zusammengefasste Evidenz aus diesen Arbeiten keine starke Leitlinienempfehlung für oder gegen die untersuchten Interventionen rechtfertigt.

2. Auf der Grundlage einer randomisierten Studie und einer systematischen Übersichtsarbeit konnten wir keinen Nutzen für eine medizinisch etablierte Intervention (Einsatz einer intraaortalen Ballongegenpulsation bei Patienten mit infarktbedingtem kardiogenen Schock) nachweisen. Die Ergebnisse dieser Arbeit führten gemeinsam mit weiteren, parallel laufenden Studien zur Änderung der Empfehlungen in evidenzbasierten europäischen und amerikanischen Leitlinien.

3. Aus der Beschreibung und Beurteilung von Verzerrungsquellen ergab sich die Frage, in welchem Ausmaß diese einen Einfluss auf den Behandlungseffekt in Studien der Intensiv- und Notfallmedizin aufweisen. Einen signifikanter Einfluss der untersuchten potentiellen Verzerrungsrisiken konnten wir für nicht nachweisen.

4. Am Beispiel von Studien aus der Intensiv- und Notfallmedizin konnten wir zeigen, dass monozentrisch- im Vergleich zu multizentrisch durchgeführten Studien zu einem im Mittel um 36 % (95 % KI 13 - 53 %) höheren Behandlungseffekt führen. Wir empfehlen deshalb zusätzliche spezifische Auswertungen für mono- und multizentrische Studien. Im Falle einer relevanten Effektmodifikation sollte die Evidenzsynthese auf der Grundlage der multizentrisch durchgeführten Studien basieren.

5. Implementierungsstrategien konnten die Leitlinienkonformität von Ärzten in der allgemeinmedizinischen Versorgung von Patienten mit Herz-Kreislauferkrankungen vor allem dann verbessern, wenn sie organisatorische Veränderungen im Versorgungsablauf, Maßnahmen zur Patienten- oder Anwender-Schulungen umfassten.

6. Eine Einbindung nicht-ärztlicher Gesundheitsberufe in die Implementierung von Leitlinien kann geeignet sein, vorhandene Barrieren gegenüber leitliniengerechten Versorgungsabläufen zu reduzieren. Die allgemeinmedizinische Versorgung von Pa-

tienten mit Herz-Kreislauferkrankungen konnte vorrangig in der Primär- und Sekundärprävention verbessert werden. Patienten leiden in der Tertiärprävention bei fortgeschrittenen Erkrankungen sowie zunehmendem Alter häufig unter komplexen Komorbiditäten. Wir betonen die Notwendigkeit, Therapieeffekte aus randomisierten Studien auf Alltagsbedingungen in der Tertiärprävention von Herz-Kreislauferkrankungen zu untersuchen.

7. Methodische Faktoren, welche den Erfolg der untersuchten Implementierungsstrategie beeinflussen können, sehen wir in der Wahl des Hauptzielkriteriums, der Durchführung der Studie mit einem individuell- oder cluster-randomisiertem Design, der Länge der Nachbeobachtungszeit und in der Kontrolle von potentiellen Verzerrungsrisiken.

9. Anlagen

Unverzagt S, Buerke M, de Waha A, Haerting J, Pietzner D, et al. (2015) Intra-aortic balloon pump counterpulsation (IABP) for myocardial infarction complicated by cardiogenic shock. Cochrane Database Syst Rev 3.

Unverzagt S, Prondzinsky R, Peinemann F (2013) Single-center trials tend to provide larger treatment effects than multicenter trials: a systematic review. J Clin Epidemiol 66: 1271-80.

Unverzagt S, Oemler M, Braun K, Klement A (2014) Strategies for guideline implementation in primary care focusing on patients with cardiovascular disease: a systematic review. Fam Pract 31: 247–66.

Unverzagt S, Peinemann F, Oemler M, Braun K, Klement A (2014) Meta-regression analyses to explain statistical heterogeneity in a systematic review of strategies for guideline implementation in primary care. Plos One 9: e110619.

Unverzagt S, Buerke M, de Waha A, Haerting J, Pietzner D, et al. (2015) Intra-aortic balloon pump counterpulsation (IABP) for myocardial infarction complicated by cardiogenic shock. Cochrane Database Syst Rev 3.

Intra-aortic balloon pump counterpulsation (IABP) for myocardial infarction complicated by cardiogenic shock (Review)

Unverzagt S, Buerke M, de Waha A, Haerting J, Pietzner D, Seyfarth M, Thiele H, Werdan K, Zeymer U, Prondzinsky R

This is a reprint of a Cochrane review, prepared and maintained by The Cochrane Collaboration and published in *The Cochrane Library* 2015, Issue 3

http://www.thecochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

Intra-aortic balloon pump counterpulsation (IABP) for myocardial infarction complicated by cardiogenic shock

Susanne Unverzagt[1], Michael Buerke[2], Antoinette de Waha[3], Johannes Haerting[1], Diana Pietzner[4], Melchior Seyfarth[5], Holger Thiele[6], Karl Werdan[2], Uwe Zeymer[7], Roland Prondzinsky[8]

[1]Institute of Medical Epidemiology, Biostatistics and Informatics, Martin Luther University Halle-Wittenberg, Halle/Saale, Germany. [2]Internal Medicine III, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany. [3]Deutsches Herzzentrum München, Department of Cardiology, Interventional Cardiology Group, München, Germany. [4]Imebi, Martin Luther University Halle, Halle, Germany. [5]HELIOS Klinikum Wuppertal, Department of Cardiology, University of Witten-Herdecke, München, Germany. [6]Medical Clinic II (Kardiology, Angiology, Intensive Care Medicine), University Clinic Schleswig-Holstein, Campus Lübeck, Lubeck, Germany. [7]Medizinische Klinik II, Klinikum Ludwigshafen, Ludwigshafen, Germany. [8]Cardiology/Intensive Care Medicine, Carl von Basedow Klinikum Merseburg, Merseburg, Germany

Contact address: Roland Prondzinsky, Cardiology/Intensive Care Medicine, Carl von Basedow Klinikum Merseburg, Weisse Mauer 42, Merseburg, 06217, Germany. r.prondzinsky@klinikum-saalekreis.de.

Editorial group: Cochrane Heart Group.
Publication status and date: New search for studies and content updated (no change to conclusions), published in Issue 3, 2015.
Review content assessed as up-to-date: 2 October 2013.

Citation: Unverzagt S, Buerke M, de Waha A, Haerting J, Pietzner D, Seyfarth M, Thiele H, Werdan K, Zeymer U, Prondzinsky R. Intra-aortic balloon pump counterpulsation (IABP) for myocardial infarction complicated by cardiogenic shock. *Cochrane Database of Systematic Reviews* 2015, Issue 3. Art. No.: CD007398. DOI: 10.1002/14651858.CD007398.pub3.

ABSTRACT

Background

Intra-aortic balloon pump counterpulsation (IABP) is currently the most commonly used mechanical assist device for patients with cardiogenic shock due to acute myocardial infarction. Although there has been only limited evidence from randomised controlled trials, the previous guidelines of the American Heart Association/American College of Cardiology (AHA/ACC) and the European Society of Cardiology (ESC) strongly recommended the use of the IABP in patients with infarction-related cardiogenic shock on the basis of pathophysiological considerations, non-randomised trials and registry data. The recent guidelines downgraded the recommendation based on a meta-analysis which could only include non-randomised trials showing conflicting results. Up to now, there have been no guideline recommendations and no actual meta-analysis including the results of the large randomised multicentre IABP-SHOCK II Trial which showed no survival benefit with IABP support. This systematic review is an update of the review published in 2011.

Objectives

To evaluate, in terms of efficacy and safety, the effect of IABP versus non-IABP or other assist devices guideline compliant standard therapy on mortality and morbidity in patients with acute myocardial infarction complicated by cardiogenic shock.

Search methods

Searches of CENTRAL, MEDLINE (Ovid) and EMBASE (Ovid), LILACS, IndMed and KoreaMed, registers of ongoing trials and proceedings of conferences were updated in October 2013. Reference lists were scanned and experts in the field contacted to obtain further information. No language restrictions were applied.

Selection criteria

Randomised controlled trials on patients with acute myocardial infarction complicated by cardiogenic shock.

Data collection and analysis

Data collection and analysis were performed according to the published protocol. Individual patient data were provided for six trials and merged with aggregate data. Summary statistics for the primary endpoints were hazard ratios (HRs) and odds ratios (ORs) with 95% confidence intervals (CIs).

Main results

Seven eligible studies were identified from a total of 2314 references. One new study with 600 patients was added to the original review. Four trials compared IABP to standard treatment and three to other percutaneous left assist devices (LVAD). Data from a total of 790 patients with acute myocardial infarction and cardiogenic shock were included in the updated meta-analysis: 406 patients were treated with IABP and 384 patients served as controls; 339 patients were treated without assisting devices and 45 patients with other LVAD. The HR for all-cause 30-day mortality of 0.95 (95% CI 0.76 to 1.19) provided no evidence for a survival benefit. Different non-fatal cardiovascular events were reported in five trials. During hospitalisation, 11 and 4 out of 364 patients from the intervention groups suffered from reinfarction or stroke, respectively. Altogether 5 out of 363 patients from the control group suffered from reinfarction or stroke. Reocclusion was treated with subsequent re-revascularization in 6 out of 352 patients from the intervention group and 13 out of 353 patients of the control group. The high incidence of complications such as moderate and severe bleeding or infection in the control groups has to be attributed to interventions with other LVAD. Possible reasons for bias were more frequent in small studies with high cross-over rates, early stopping and the inclusion of patients with IABP at randomisation.

Authors' conclusions

Available evidence suggests that IABP may have a beneficial effect on some haemodynamic parameters. However, this did not result in survival benefits so there is no convincing randomised data to support the use of IABP in infarct-related cardiogenic shock.

PLAIN LANGUAGE SUMMARY

Intra-aortic balloon counterpulsation in patients with acute myocardial infarction and cardiogenic shock

Cardiogenic shock is a severe condition in which a suddenly weakened heart is not able to pump enough blood to meet the body's energy needs, so not enough oxygen will reach the body's organs. Cardiogenic shock is a life-threatening medical emergency and needs to be treated quickly to avoid organ damage or even death of the affected patient. Most often cardiogenic shock is caused by a severe heart attack and the induced damage to the heart muscle. Despite more than 50 years of effort, patients with cardiogenic shock still have a poor prognosis after primary revascularization procedures such as coronary artery bypass grafting or primary percutaneous coronary intervention. The main cause for the development of cardiogenic shock is the loss of myocardial function due to myocardial infarction leading to impaired left ventricular function with unstable haemodynamics and reduced systolic and mean arterial pressures. The reduced blood pressure leads to hypoperfusion and so reduced oxygen supply to vital organs and the corresponding clinical signs. These include cold and pale skin, reduced or a lack of urine output and signs of impaired cerebral function like dizziness or even unconsciousness.

On this basis, it was reasoned that the use of mechanical means of augmenting pressure and flow would prove effective. The very first mechanical means of assisting the circulation in such a manner was by a counter pulsation strategy using a device called the intra-aortic balloon pump (IABP). Through balloon inflations and deflations synchronized with the natural heartbeat the IABP increases diastolic aortic pressure, which enhances diastolic blood flow to the coronary arteries and vital organs, as well as reduces systolic aortic pressure, which reduces afterload and oxygen consumption of the myocardium and increases cardiac output. This support can be provided for a few hours and, in extreme cases, for several weeks. Evidence from earlier published studies suggested that certain patients with acute myocardial infarction complicated by cardiogenic shock and treated by thrombolysis may derive benefit from a period of support with the IABP. However, nowadays the most widely recommended and preferred revascularization procedure is primary percutaneous coronary intervention.

In contrast to the previous version of this review, this update now includes data from one large and six small randomised controlled trials. It allows more definitive conclusions about the potential beneficial or harmful clinical effects of IABP support beyond its immediate haemodynamic effects. Complications such as moderate and severe bleeding were more frequently observed in patients treated with more invasive devices than IABP. Small randomised trials suffered from inadequate power to address deaths and harmful effects of IABP and were biased by frequent cross-over to the more aggressive strategy, early stopping of the trial, or the inclusion of patients with IABP at randomisation. It is most noteworthy that a recently conducted and published large randomised trial showed no evidence for survival

benefits of IABP support in patients with infarct-related cardiogenic shock treated by percutaneous coronary intervention (PCI). On the basis of these data, IABP support is no longer strongly recommended by the European Society of Cardiology (ESC) guidelines for treatment of patients with infarct-related cardiogenic shock. Rather, IABP use is based on the personal experience and decision of the physician and the particular circumstances of individual patients.

BACKGROUND

Description of the condition

Worldwide, cardiovascular disease is estimated to be the leading cause of death and loss of disability-adjusted life years (Gaziano 2010; Lozano 2012; Moran 2014; Mozaffarian 2014; Nieuwlaat 2013; The Global Burden Collaboration 2014). Each year approximately 920,000 people in the United States (US) experience acute myocardial infarction (AMI), with a prevalence of 5.1% of all males and 2.5% of all females over 20 years old and about 150,000 of them die. The estimated direct and indirect 2008 costs of coronary heart disease (ICD 10 codes I20 to I25) in the US was USD 156.4 billion (AHA 2008). In the United Kingdom about 227,000 myocardial infarctions occur annually and it has been estimated that about 1 million people over 35 years old have had a myocardial infarction (BHF 2007). Data from the INTERHEART study showed that the rates of cardiovascular disease have risen greatly in low-income and middle-income countries, with about 80% of the global burden of cardiovascular disease occurring in these countries (Yusuf 2004).

AMI is complicated by cardiogenic shock in 7% to 10% of cases (Goldberg 1999; Hochman 1999). Cardiogenic shock after AMI is a complex syndrome that involves a cascade of acute left ventricular dysfunction, decreased cardiac output, hypotension and tissue hypoperfusion (Hochman 2007). Subsequently, complicating multi-organ dysfunction might occur due to ischaemia and reperfusion and the following inflammatory response. Clinically defined, cardiogenic shock is hypotension (a systolic blood pressure of < 90 mmHg for at least 30 minutes or the need for supportive measures to maintain a systolic blood pressure of ≥ 90 mmHg) and end-organ hypoperfusion (cool extremities or a urine output of < 30 mL per hour, altered mental status, or elevated serum lactate). Haemodynamic criteria that are sometimes used include cardiac index (< 1.8 L/min/m^2 or < 2.2 L/min/m^2 if inotropic drugs or vasopressors are used) and a pulmonary capillary wedge pressure of at least 15 mmHg (Forrester 1976a; Forrester 1976b; Hochman 1999). Patients with sustained hypotension, suspected cardiogenic shock or suspected acute heart failure at the time of AMI are at increased risk of death, approaching 30% to 70% mortality within 30 days (Ohman 2005a; Thiele 2012; Werdan 2014). Fewer than 50% of patients with cardiogenic shock survive up to one year (Hochman 2007).

The poor outcome associated with medical management of cardiogenic shock has spurred more aggressive interventional approaches, including thrombolysis, intra-aortic balloon pump counterpulsation (IABP) support and early diagnostic angiography with primary percutaneous coronary revascularization (Ohman 2005a). Early mechanical revascularization, using either percutaneous coronary intervention (PCI) or coronary artery bypass graft (CABG) surgery, along with supportive care, improves mid- and long-term survival in these patients when compared with initial medical stabilisation alone (Hochman 2001).

Description of the intervention

IABP was introduced into clinical practice in 1968 (Kantrowitz 1968) as a means for supporting patients undergoing surgical revascularization. Initial experience documented that this device had important physiological effects including an improvement of cardiac function and diastolic blood pressure and a reduction in systemic acidosis. More recently, in registry trials, several investigators have shown enhanced coronary, cerebral and renal perfusion (even microperfusion) with IABP, particularly among patients having PCI during cardiogenic shock. However, this impressive physiological profile has not been followed by equally important randomised clinical trial data (Ohman 2001; Werdan 2010).

During the last few years IABP insertion has become safer with the smaller diameters of the balloon catheters and the corresponding insertion sheaths. Nevertheless every additional arterial puncture, especially in the emergency situation, can lead to IABP-related complications such as bleeding, arterial ischaemia, venous thrombosis and also infections. Because bleeding complications, in particular, are not always defined and reported according to comparable standards, the evaluation of IABP-related complications remains hampered. Studies reporting complication rates are diverse in terms of the indications for aortic counterpulsation, the technique used for insertion (surgical or percutaneous), the duration of use and the specific definition of the complication itself (Arafa 1999; Assis 2009; Cohen 2003; Cooper 2008; Dyub 2008; Erdogan 2006; Fuchs 2009; Gjesdal 2009; Kocogullari 2008; Kumbasar 1999; Lewis 2007; Pfeiffer 2005; Riaz 2008;

Stone 2003). The presence of peripheral arterial disease (including a history of claudication, femoral bruit, or absent pulses) has been the most consistent and reproducible predictor of complications (Santa-Cruz 2006). Arafa 1999 reported major vascular complications (limb ischaemia, aortic dissection, abdominal aorta perforation, bilateral limb ischaemia) in 8%, minor vascular complications (haematoma requiring operative revision, haemorrhage treated by IABP removal, limb ischaemia relieved by IABP removal, local infection and ischaemic skin loss) in 3%, and late vascular complications in 2% of patients (foot drop, pseudoaneurysm, limb ischaemia). Newer platelet inhibitors, such as prasugrel and ticagrelor, and anticoagulant drugs like the direct factor Xa- and thrombin-inhibitors are increasingly being used and may lead to increased bleeding risks with IABP support.

Why it is important to do this review

IABP has been the most commonly used mechanical assist device for patients with cardiogenic shock for more than four decades. Its use has been encouraged by a class I recommendation in the previous American Heart Association (AHA)/American College of Cardiology (ACC) and also the European Society of Cardiology (ESC) guidelines for the management of AMI patients with cardiogenic shock (Antman 2004; van de Werf 2008). The Level B in the AHA/ACC and Level C evidence in the ESC guidelines supporting this recommendation could largely be attributed to pathophysiological considerations and benefits observed in registries that predominantly enrolled patients treated with thrombolytic therapy in the pre-PCI era. There have been controversial differences in therapeutic behaviour in the US and European countries, with far greater use of IABP in the US than in Europe. In the US, with the highest rate of IABP use, the mortality rate was lower than in European countries such as the United Kingdom (Hudson 1999).

In the early 1980s two smaller randomised trials failed to show any benefit of IABP, compared with control therapy, on infarct size or left ventricular function in patients with myocardial infarction predominantly without cardiogenic shock (Flaherty 1985; O'Rourke 1981). Previous randomised trials of IABP in high-risk patients without cardiogenic shock have suggested a lower morbidity, particularly among the patients with several high-risk features (Ishihara 1991; Ohman 1994). These studies were too small to address mortality but they favoured better outcomes with IABP treatment compared to standard therapy without IABP. Randomised trials of IABP in cardiogenic shock were clearly needed, and one was conducted with surrogate endpoints (Prondzinsky 2010) and another was not completed because of physician bias and difficulties in obtaining consent among critically ill patients (Ohman 2005). The IABP SHOCK II Trial was the first large randomised controlled multicentre trial adequately powered to investigate the influence of IABP support on mortality, and enrolled 600 patients with infarct-related cardiogenic shock (Thiele 2012).

Non-randomised clinical studies have nearly uniformly shown a benefit associated with IABP for patients with cardiogenic shock (Alcan 1983; Forssell 1979; Holmes 1997; Kontovannis 1999; Kovack 1997b; McEnany 1978; Moulopoulos 1986b; Takano 1984; Weiss 1984). However, these studies are subject to selection bias and patients receiving IABP were in general younger, had fewer comorbid illnesses, and were more aggressively treated with cardiac catheterization and revascularization compared with patients not treated with IABP (Hudson 1999; Sanborn 2000). Data from a large prospective registry suggest little benefit of IABP placement in cardiogenic shock patients treated with primary PCI (Barron 2001), and one trial reported higher mortality rates associated with IABP use in this group of patients (Barron 2001).

It seems probable that not all patients in cardiogenic shock benefit from IABP therapy (Hochman 2003), in particular in those where the component of inflammation was associated with systemic inflammatory response syndrome and septic organ failure. This begs the question whether IABP may be beneficial in inflammatory conditions and whether IABP may accelerate systemic inflammation by continuous blood cell surface activation.

A systematic review by Theologou 2011 suggests that preoperative IABP use may be beneficial on mortality and morbidity in specific high-risk patients groups undergoing coronary artery bypass surgery. However, they state many problems with the quality, validity and generalisability of the trials.

A systematic review by Sjauw 2009 of IABP therapy in ST-elevation myocardial infarction (STEMI) performed two separate meta-analyses. The first meta-analysis included seven randomised trials in 1009 patients with STEMI restricted to patients without cardiogenic shock and the second used data from non-randomised trials of 10,529 STEMI patients with cardiogenic shock. A second systematic review by Cheng 2009 performed a meta-analysis of three studies comparing the safety and efficacy of IABP with percutaneous left ventricular assist devices (LVADs) and performed a meta-analysis of aggregate data for 30-day survival, haemodynamics (cardiac index (CI), mean arterial blood pressure (MAP) and pulmonary capillary wedge pressure (PCWP)) and adverse events (leg ischaemia, bleeding and sepsis).

Up to now no systematic review integrating the data of the IABP SHOCK II Trial has been performed and published. The results of this updated review will add more evidence and this review provides a formal assessment of the cumulative data with meta-analysis of all the evidence for and against the use of IABP in patients with acute myocardial infarction complicated by cardiogenic shock. IABP insertion in critically ill patients is correlated with the risk of complications and these potential risks can only be justified by an acceptable (evidence-based) assessment of measurable beneficial clinical effects in IABP-treated patients. This will have implications for clinical practice. Three additional studies comparing IABP versus standard treatment without IABP (Arias 2005; Prondzinsky 2010; Thiele 2012) and subgroups of patients with myocardial infarction and cardiogenic shock from two other

studies (Burkhoff 2006; Ohman 2005) were included. Extensive analyses of 30-day, 6-month and 12-month mortality distributions provide an important opportunity to examine the effects over a prolonged period of time. Analyses were adjusted for age, sex and diabetes as a comorbidity to see whether the observed effects were consistent across different types of patients.
As a consequence of the data from one systematic review (Sjauw 2009) the corresponding guidelines of ACC/AHA and ESC, and also the German-Austrian guideline, for the treatment of infarct-related cardiogenic shock have been revised (O'Gara 2013; Steg 2012; Werdan 2012).

OBJECTIVES

The primary aim of this review was to evaluate, in terms of efficacy and safety, the effect of IABP versus non-IABP or other assist device guideline compliant standard therapy on mortality and morbidity in patients with AMI complicated by cardiogenic shock.

METHODS

Criteria for considering studies for this review

Types of studies

Randomised controlled trials with or without blinding and any report of mortality that examined the efficacy of IABP versus standard therapy were included. We accepted studies with cross-over of individual patients. Observational studies were excluded.

Types of participants

Adult patients (from the age of 18 years) with a clinical diagnosis of myocardial infarction complicated by cardiogenic shock undergoing PCI, coronary artery bypass bypass graft surgery (CABG) or thrombolysis.

Types of interventions

IABP versus non-IABP or other assist device guideline compliant standard therapy. The term standard therapy describes guideline compliant therapies (PCI, CABG, surgery or thrombolysis, pharmacological haemodynamic and, as required, ventilatory or other organ function support).

Types of outcome measures

Primary outcomes

- All-cause mortality (mortality distribution and rates within the commonly accepted limits, either to discharge, within 30 days, 6 months and 1 year)
- Non-fatal cardiovascular events (reinfarction, reocclusion and subsequent re-revascularization, stroke, recurrent ischaemia) (hierarchical lower ranked endpoint)

Secondary outcomes

- Haemodynamics (cardiac index (CI), mean arterial blood pressure (MAP), pulmonary capillary wedge pressure (PCWP))
- Length of hospital and intensive care unit (ICU) stay
- Quality of life
- All IABP-related post-interventional complications

Search methods for identification of studies

Searches were conducted to identify published and unpublished randomised controlled trials. Searching for trials included all information available since 1968 (introduction of IABP into clinical practice (Kantrowitz 1968)) up to October 2013. No language restrictions were included in the search strategies.

Electronic searches

The search strategies for the review were constructed by using a combination of subject headings and terms relating to the health condition of interest (myocardial infarction and cardiogenic shock), the intervention (intra-aortic balloon pump counterpulsation) and the type of study design (randomised controlled trial). We used controlled vocabulary terms and text words and searched different sources. The search strategies used are documented in Appendix 1 (2010) and Appendix 2 (2013).
The following sources were searched. Health-related electronic bibliographic databases: the Cochrane Central Register of Controlled Trials (CENTRAL) in *The Cochrane Library* (Issue 9, 2013), MEDLINE (Ovid) (1946 to September 2013, searched 2 October 2013), EMBASE (1947 to September 2013, searched 2 October 2013), PubMed (searched 2 October 2013), LILACS, IndMed and KoreaMed (unrestricted date to November 2013, searched 12 November 2013).

Searching other resources

The following sources were searched.
Registers of ongoing and completed trials:

- www.controlled-trials.com (searched 12 November 2013);
- www.centerwatch.com (searched 12 November 2013);

- Benchmark registry (www.datascope.com/ca/pdf/benchmark2_brochure.pdf); and
- National Research Register (www.controlled-trials.com/mrt/archived (2000 to 2007).

Handsearching included the annual conference proceedings of the following societies: American Heart Association (AHA) (published in Circulation), American College of Cardiology (ACC), European Society of Cardiology (ESC), European Society of Intensive Care (ESICM) and Deutsche Gesellschaft für Kardiologie (all 1968 to 2013).
Members of the Cochrane Heart Group, experts in the field, and manufacturers of the device were contacted. In addition, reference lists from eligible trials were scanned and first authors were contacted to obtain further information on study design and to collect individual patient data.

Data collection and analysis

Selection of studies

Studies identified through the search strategies described above were screened by the titles. In a second step, two authors (SU, RP or MM) independently screened abstracts and keywords. Full-text articles were taken into account for further assessment if the information given suggested that the study:

- used random or quasi-random allocation to the comparison groups (IABP versus non-IABP);
- included patients with myocardial infarction complicated by cardiogenic shock;
- included primary data.

Differences in opinion were settled by consensus with a third review author. After the exclusion of non-relevant publications and duplicates, the full-text versions of the remaining papers were assessed against the inclusion and exclusion criteria and data were extracted and entered into standardised data extraction forms.The selection process was recorded in a PRISMA flow chart (Moher 2009).

Data extraction and management

Two authors (SU, RP or MM) independently extracted details of study population, interventions and outcomes by using a data extraction form, which was designed especially for this review. Differences in data extraction were resolved by consensus with a third author, and referring back to the original article. The data extraction form included the following items.

- General information: title, authors, source, contact address, country, published or unpublished, language and year of publication, trial sponsor.
- Trial characteristics: including study design, timing and follow-up, quality assessment as specified above.
- Patients: inclusion and exclusion criteria, sample size, baseline characteristics, similarity of groups at baseline, withdrawals, cross-over, losses to follow-up.
- Interventions: type of standard therapy and comparison assist devices.
- Outcomes: time to death (hazard ratios (HR) and their 95% confidence intervals (CI)), number of deaths and patients per group, mortality at specific time points (in-hospital, at 30 days, 6 months, 1 year), other clinical event outcomes (reinfarction, reocclusion, re-revascularization, stroke, recurrent ischaemia), haemodynamics (CI, MAP, PCWP), length of hospital and ICU stay, doses of catecholamines (dobutamine, norepinephrine, dopamine), IABP-related post-interventional complications.

As this review was planned as an individual patient data (IPD) meta-analysis, first authors of all eligible trials were contacted and asked to provide IPD and data on missing information. Analyses are based on updated IPD as a reliable and the most powerful method to calculate and compare times to death with adjustments for important co-variables (age, sex and diabetes) (Piedbois 2004).

Assessment of risk of bias in included studies

The review analyses the results of randomised controlled trials (RCTs). Two authors (SU, RP or MM) independently assessed the internal validity of eligible studies according to the Cochrane Collaboration risk of bias tool (Higgins 2011). Disagreements were resolved in discussion with RP and HT until consensus was obtained.
Risk of bias was described and judged in six specific domains:

- sequence generation;
- allocation concealment;
- blinding of participants, personnel, and outcome assessors;
- incomplete outcome data;
- selective outcome reporting; and
- other sources of bias (such as cross-over of patients, early stopping, per protocol analysis).

The domains of sequence generation, allocation concealment and selective outcome reporting were reported by a single entry for each study. For incomplete outcome data two entries were used because assessments generally need to be made separately for different outcomes (mortality and haemodynamics). Blinding of the investigated intervention was judged to be not possible. The description was based on the published study report, which was added to by a mixture of study reports, protocols, published comments on the study and contact with the investigators.

Measures of treatment effect

Meta-analysis was conducted on mortality distribution and mortality rates, non-fatal events, haemodynamics (CI, MAP, PCWP)

measured within six hours after implantation and length of hospital and ICU stay on the basis of individual patient or published aggregate data.

The HR and 95% CI for time to death as one of the primary outcome measures were calculated for seven studies with available IPD to describe the mortality distribution over 30 days and over the whole investigation period. For these trials, Kaplan-Meier curves and mortality rates at discharge from hospital and at 30 days, six months and one year after randomisation were generated. Mortality rates of all eligible trials were compared and odds ratios were calculated. Because of high survival rates, it was not possible to calculate median survival for both treatment groups in one of the included trials. Weighted mean differences (WMDs) were calculated as effect measures for haemodynamics and length of hospital stay, and odds ratios (ORs) for IABP-related complications. All final effect measures were presented with their 95% CI.

Safety outcomes to describe IABP-related post-interventional complications were chosen a posteriori and included the following frequently reported possible device-related adverse events during support: bleeding, vascular injury, leg or limb ischaemia, embolism, infection and thrombocytopenia. The frequencies of these IABP-related complications and non-fatal cardiovascular events were presented in numbers and percentages with corresponding ORs.

Unit of analysis issues

Patients were individually randomised into two groups. Studies with cross-over of individual patients were included. These patients were investigated in their randomisation group according to the intention-to-treat (ITT) principle. A per protocol (PP) analysis was added in studies where the overall proportion of cross-over compared with the observed event risk might have had a clinically relevant impact on the intervention effect, and available information on cross-over and all-cause mortality (Other potential sources of bias). The effect of the intervention was measured and analysed on the basis of single measurements for each outcome for each patient.

Dealing with missing data

If data were not available in the trial report or data collection the investigators were approached to see if the missing data could be provided. Only in one RCT (Arias 2005) HRs and 95% CIs were not calculated because of missing information. The first author was contacted and provided some of the missing information. He was not able to provide any individual patient data and had no access to the database. ORs were used to describe the effect on in-hospital mortality in this trial. All other HRs were calculated from individual patient data.

Assessment of heterogeneity

Heterogeneity was classified by two independent review authors on methodological and clinical grounds. Inconsistency between studies was quantified by the I^2 statistic (Higgins 2002). In the case of substantial clinical, methodological or statistical heterogeneity ($I^2 > 50\%$) meta-analysis was restricted to subgroups. Independent of the presence of statistical heterogeneity, possible causes were assessed if the differences in outcomes seemed clinically important.

Assessment of reporting biases

Although every effort was made to identify unpublished studies, publication bias was assessed.

Data synthesis

The analysis was based on the intention-to-treat (ITT) principle. The IPD analysis contained data from all randomised patients with AMI and cardiogenic shock from six of the seven relevant studies. Analyses of IPD were done using SAS software (Whitehead 2002). First, all trials were analysed individually and finally a stratified Cox model of all trials with different baseline hazard functions in each single trial was used to estimate the overall HR (one-step approach). Based on the high heterogeneity between the included RCTs (differences in the treatments in the control groups, in pharmacological support, length of follow-up, primary outcome measures, sources of bias) we decided to use the random-effects model for meta-analysis of the relevant studies. The one-step meta-analysis as described above and a two-step approach (Riley 2010) with separate Cox models in single trials and data synthesis in RevMan gave nearly identical results. We show the results of the one-step meta-analyses to describe all-cause mortality distribution in the text and all additional analyses in Table 1. Data and analysis tables and forest plots display effect estimates and CIs for both individual studies and the two-step meta-analyses. IPD were not provided for one study. We reduced our available IPD describing in-hospital-mortality rates to aggregated data and combined the aggregate data by the two-stage-approach described in Riley 2007. Non-fatal cardiovascular events, IABP-related post-interventional complications and all secondary outcome measures were analysed descriptively with RevMan 5.2.

Dealing with the proportional hazards assumption

The elementary assumption in the Cox proportional hazards model demands a constant effect (or a constant quotient of the hazards over the observation period). Most study data did not satisfy this assumption.To test the extent of this infraction a gamma frailty model (Duchateau 2008) was applied. This analysis showed only a weak influence on the parameter estimation and we proceeded to execute the analysis as preplanned with the Cox proportional hazards model.

Subgroup analysis and investigation of heterogeneity

Stratified analyses were restricted to preplanned prognostic factors: age (< 75 versus ≥ 75 years), diabetes and sex to find differences in survival with IABP support.

Sensitivity analysis

Sensitivity analyses were performed to explore the influence of including or excluding certain types of studies. Due to the low number of heterogeneous included studies and different sources of bias we restricted our analyses to the preplanned influence of standard therapy (PCI versus thrombolytic therapy) and added a sensitivity analysis to investigate the influence of different types of controls (with or without other LVAD).

RESULTS

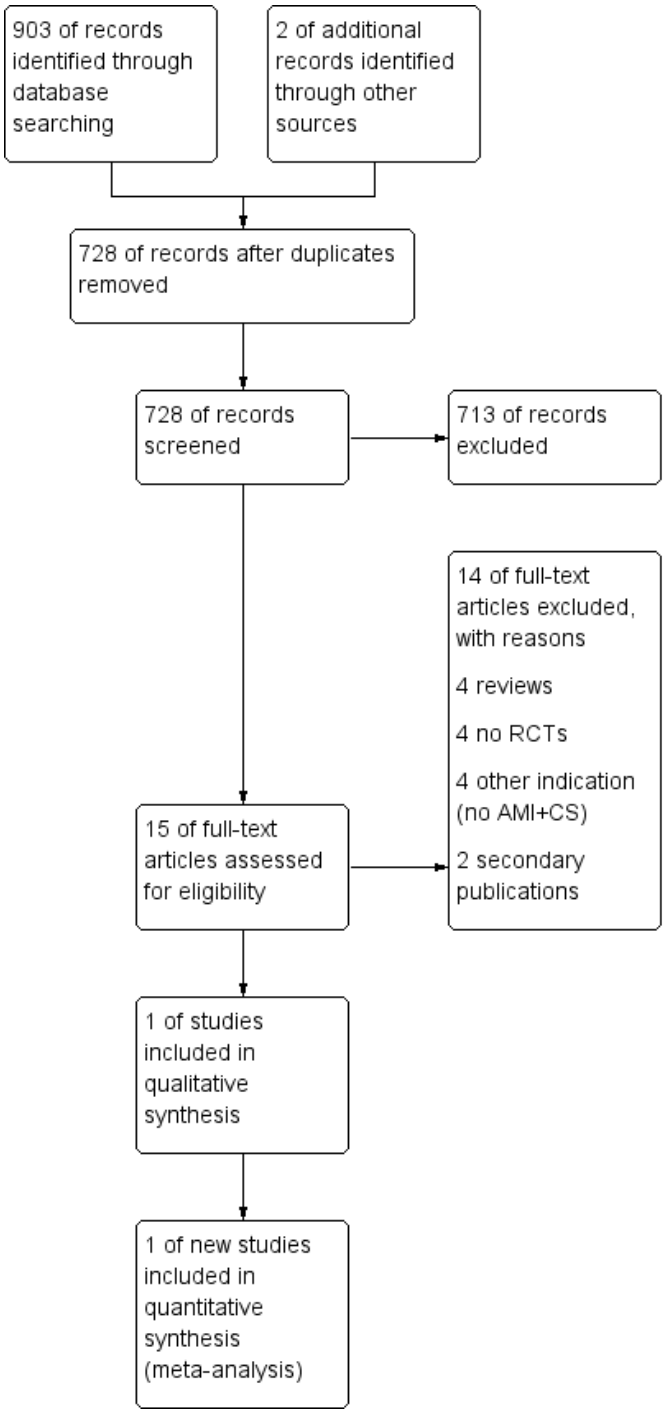
Description of studies

Results of the search

Having used the search strategies (Appendix 1) in January 2010, a total of 1410 potentially relevant references were identified (CENTRAL 5, MEDLINE 757, EMBASE 639 and other 9). The update in October 2013 identified 904 references (CENTRAL 126, MEDLINE 386, EMBASE 376, PubMed 15 and other 1) (search strategies in Appendix 2).

Forty-eight studies were thought to be of relevance and full papers were assessed against the inclusion and exclusion criteria. Of these only seven met our predefined inclusion criteria. The remaining studies are listed in Characteristics of excluded studies. This update search process was recorded in PRISMA flow charts (Figure 1).

Figure 1. Flow diagram update January 2013.



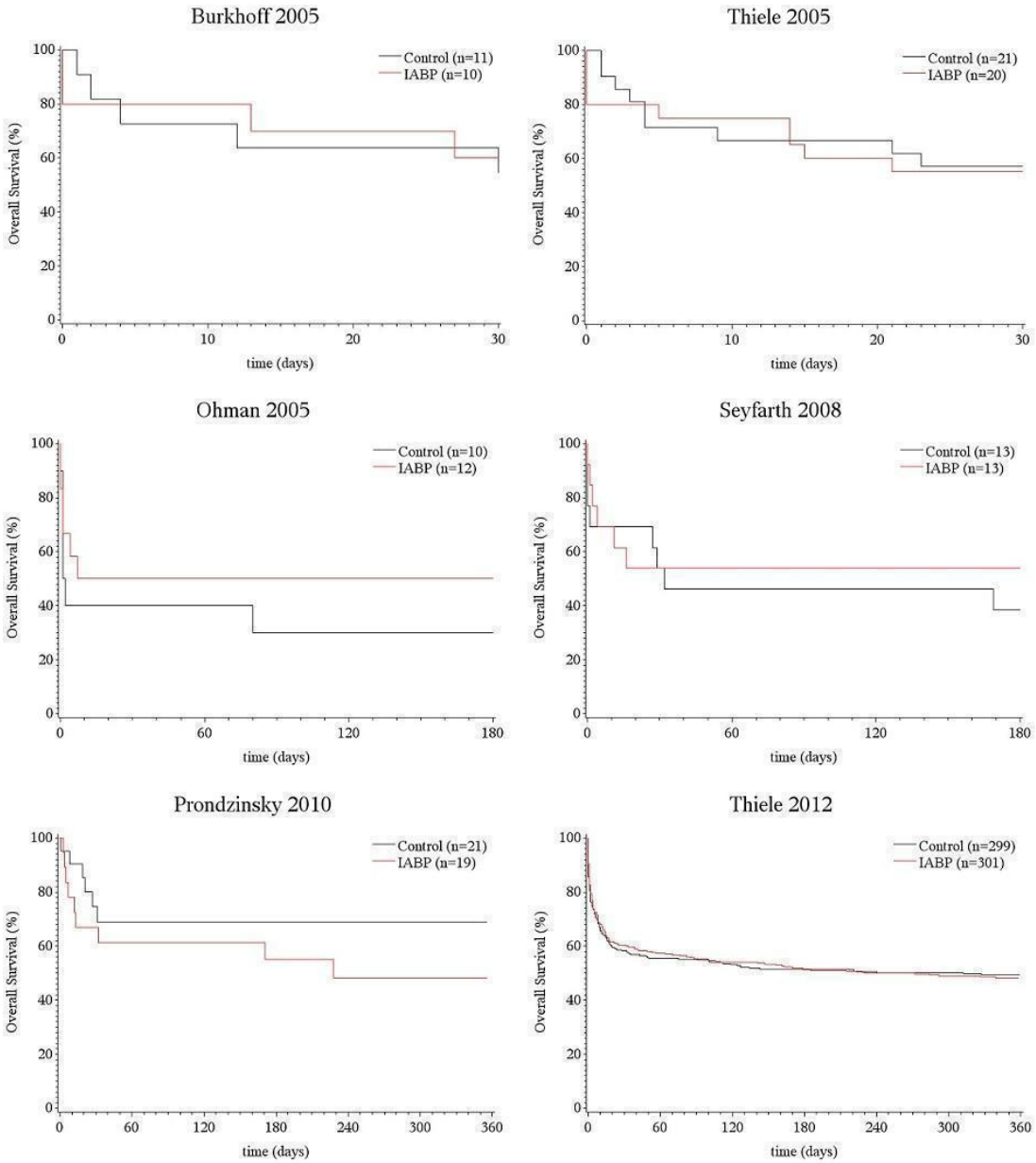
Included studies

Seven eligible studies with a total of 790 patients with AMI and cardiogenic shock were identified for the comparison IABP versus no IABP (Arias 2005; Burkhoff 2006; Ohman 2005; Prondzinsky 2010; Seyfarth 2008; Thiele 2005; Thiele 2012). Four studies were conducted in Germany (Prondzinsky 2010; Seyfarth 2008; Thiele 2005; Thiele 2012), one study in the United States (Burkhoff 2006), one study in Mexico (Arias 2005) and one study in the United States, Australia and Europe (Ohman 2005). Most patients had Caucasian or Hispanic ethnicity.

Authors of all included studies provided additional information. Individual patient data (IPD) from six studies with 750 patients were collected in the database. Each study characteristic is presented briefly in tabulated form (please see Characteristics of included studies and Characteristics of ongoing studies). A more comprehensive assessment of the included studies is given below.

Ohman 2005 randomised between 1996 and 1999 57 patients in a multicentre open label trial into two groups, a control group (27) and an intervention group (30). Analysis and descriptions were restricted to 22 patients with suspected cardiogenic shock in Killip class IV (12 in the intervention group and 10 in the control group). The intervention group got IABP up to three hours after starting fibrinolysis, by the standard or sheathless technique. Patients received IABP for 48 hours at a rate of 1:1 and were weaned gradually over 12 hours before pump removal. In total 3/10 (33%) of patients with cardiogenic shock from the control group crossed over to emergency IABP, and 3/12 (25%) of patients from the intervention group did not receive IABP. The mean duration of support was 53 ± 30 hours in the intervention group. Myocardial revascularization was performed using fibrinolytic therapy (all patients), PCI (23%), stent implantation (14%) or bypass surgery (18%). Pharmacological support with intravenous heparin was pre-specified, and use of other medications and procedures was left to the discretion of the physicians. During the first 30 days 6/12 patients (50%) died in the intervention group and 6/10 patients (60%) in the control group (OR 0.67, 95% CI 0.12 to 3.64). Six months after randomisation a total of 6/11 patients (55%) in the intervention group and 7/10 patients (70%) in the control group had died (OR 0.51, 95%CI 0.09 to 3.11). The resulting HRs of all-cause 30-day mortality distribution (HR 0.75, 95% CI 0.28 to 2.00) and 6-month mortality distribution (HR 0.66, 95% CI 0.22 to 1.99, log-rank P = 0.42) (Figure 2) slightly favoured the intervention with IABP but did not reach statistical significance. Adjustments for age, sex and diabetes showed little influence on these results; 2/12 patients (17%) suffered from stroke in the intervention group and no reinfarctions were documented as non-fatal cardiovascular events during hospital stay. There was only one related complication in the control group (limb ischaemia). Only clinical endpoints and no haemodynamic parameters or information about in-hospital stay and intensive care requirement were provided. The trial was stopped early because it did not achieve the enrolment goal.

Figure 2. All-cause mortality distribution of included studies on the basis of individual patient data



Thiele 2005 between 2000 and 2003 randomised 41 patients in a single-centre open label trial into two groups, a control group (21) and an intervention group (20). The inclusion criteria were myocardial infarction complicated by cardiogenic shock. The intervention group got IABP percutaneously according to standard procedures (initially on a pumping ratio of 1:1 with 100% balloon inflation); the control group got treatment with a percutaneous LVAD (TandemHeart). One patient with rapid haemodynamic improvement did not receive an LVAD. Mean duration of support with IABP was 84 ± 54 hours in the intervention group and 77 ± 47 hours with TandemHeart in the control group. Myocardial revascularization was performed using PCI (plus stenting) (95% of patients) and bypass surgery (5% of patients). Pharmacological support on the basis of heparin, dopamine and dobutamine, diuretics and fluids was given according to standard intensive care guidelines. All patients with PCI were started with aspirin and clopidogrel. Mortality during the first 30 days included 9/20 patients (45%) in the intervention and 9/21 patients (43%) in the control group (OR 1.09, 95% CI 0.32 to 3.75). The HR of the all-cause 30-day mortality distribution (HR 1.09, 95% CI 0.44 to 2.79, log rank P = 0.86) (Figure 2) reflected no significant difference between groups. Adjustments for age, sex and diabetes showed only little influence on these results. During hospital stay 10/20 patients from the intervention group died (50%) (in-hospital mortality rate OR 1.33, 95% CI 0.39 to 4.57). Three patients were documented with four non-fatal cardiovascular events during the in-hospital stay: one patient with reinfarction and one with recurrent ischaemia in the intervention group, and one patient with reocclusion and subsequent re-revascularization in the control group. The pre-implantation cardiac index (CI) was noted to fall in the intervention group (1.62 ± 0.37 to 1.19 ± 0.84 L/min/m^2 post-implantation, P = 0.058) and to rise in the control group (1.71 ± 0.38 to 2.32 ± 0.59 L/min/m^2 post-implantation, P < 0.001). Pre-implantation MAP was noted to rise in the intervention group (65 ± 14 to 72 ± 12 mmHg post-implantation, P = 0.003) as well as in the control group (62 ± 14 to 76 ± 13 mmHg post-implantation, P < 0.001). Pre-implantation PCWP was noted to fall in the intervention (25.1 ± 6.1 to 21.6 ± 5.8 mmHg post implantation, P = 0.028) as well as in the control group (20.8 ± 4.2 to 15.9 ± 3.8 mmHg post-implantation, P < 0.001). Length of in-hospital stay was 12 ± 14 days in the intervention group and 13 ± 13 days in the control group. There were some possibly related complications in both treatment groups. A total of 5/20 patients (25%) from the intervention group and 18/21 patients (86%) from the control group suffered from moderate or severe bleeding. No patient from the intervention group but 7/21 patients (33%) from the control group developed limb ischaemia after implantation of a 17-French arterial cannula; 12/20 patients (60%) from the intervention group and 14/21 patients (66%) from the control group suffered from infections. One patient from the intervention group suffered from embolism and one patient from the control group developed thrombocytopenia.

Arias 2005 between 2001 and 2003 randomised 40 patients into two groups in a single-centre open label trial. The authors analysed patients in a control group (9) and an intervention group (31): 27.5% of patients crossed over to IABP. The inclusion criteria were myocardial infarction complicated by cardiogenic shock. The intervention group got IABP percutaneously with fluoroscopy, an Arrow AutoCAT 2 WAVE® IABP. Myocardial revascularization was performed using PCI. Pharmacological support was given on the basis of inotropic (dopamine and dobutamine), vasopressor, analgesic and anticoagulant agents. In-hospital mortality rates from the coronary station included 10/31 deaths (32%) in the intervention group and 5/9 deaths (56%) in the control group. The resulting OR (0.38, 95% CI 0.08 to 1.73) slightly favoured intervention with IABP but did not reach statistical significance.

Burkhoff 2006 between 2002 and 2004 randomised 33 patients in a multicentre open label trial into two groups, a control group (19) and an intervention group (14). Of the 33 randomised patients, 21 were diagnosed with AMI (10 in the intervention group and 11 in the control group). Additionally, nine patients were treated in the roll-in phase, five of them with AMI: 36% of patients from the intervention group were bridged to another therapy after enrolment (four patients to LVAD, one patient to PCI) and 37% of patients from the control group were bridged to another therapy (three patients to LVAD, one patient to extracorporeal membrane oxygenation, two patients to PCI with stenting placement, one patient to mitral valve repair). The intervention group got conventional treatment with IABP, the control group got treatment with a percutaneous LVAD (TandemHeart). Most patients (67%) entered the study on IABP but still met haemodynamic criteria for cardiogenic shock. The mean duration of support with IABP was 75 ± 95 hours in the intervention group and 61 ± 45 hours with TandemHeart in the control group. Myocardial revascularization was performed for patients with myocardial infarction using PCI (85% of patients), bypass surgery (12%) or LVAD (4%). Pharmacological support on the basis of vasopressor, inotropic and pharmacologic agents was based on the physician's standard of care. Mortality of cardiogenic shock and AMI patients during the first 30 days included 4/10 patients (40%) in the intervention group and 4/11 patients (36%) in the control group (OR 1.17, 95% CI 0.20 to 6.80). The resulting HR of the all-cause 30-day mortality distribution (HR 0.88, 95% CI 0.23 to 3.31, log rank P = 0.85) (Figure 2) reflected no significant difference between groups. Adjustments for age and sex showed no influence on these results. Predefined haemodynamic success criteria (no death during support or within 24 hours of device removal, CI ≥ 2.2 L/min/m^2, PCWP ≤ 24 mmHg and MAP ≥ 70 mmHg reflecting the average values during support) were satisfied in 14% of all randomised

patients in the intervention group compared with 37% of patients in the control group. Most of the patients entered the study already on IABP and were then randomised to continued IABP or to switch to TandemHeart. Therefore, pre-IABP haemodynamic information was not available and we decided not to include the haemodynamic parameters from this trial. There were some possibly related complications in both treatment groups. In the intervention group there was one need for surgical intervention to treat a device-related adverse event (7.1%) and one device-related removal because of a problem (7.1%). In the control group there was one instance of device failure (5.3%). On average, patients in the intervention group experienced 2.6 events per patient (1.2 serious) compared with 3.1 events per patient (1.3 serious) in the control group. There were no specific adverse events related to the performance of the trans-septal puncture or insertion of the trans-septal cannula; 2/14 patients (14%) from the intervention and 8/19 patients (42%) from the control group suffered from bleeding; 2/14 patients (14%) from the intervention group and 4/19 patients (21%) from the control group developed leg ischaemia. No patient from the intervention group but 3/19 patients (16%) from the control group suffered from cannulation site infection; 3/14 patients (21%) from the intervention group and 3/19 patients (16%) from the control group suffered from thrombocytopenia. Proper discrimination between device-related and shock-induced symptoms was not performed according to the frequent occurrence of neurologic dysfunction.

Seyfarth 2008 between 2004 and 2007 randomised 26 patients in a two-centre open label trial into two groups, a control group (13) and an intervention group (13). The intervention group got conventional treatment with IABP, the control group got treatment with a percutaneous LVAD (Impella). One patient assigned to the control group died before implantation and did not receive an LVAD. The assigned device was implanted in both groups after revascularization therapy, via the access site. As long as the assigned device was implanted, heparin was given intravenously, adjusted to a partial thromboplastin time of 60 to 80 seconds. The mean duration of support with IABP was 26 ± 19 hours in the intervention group and 27 ± 16 hours with Impella in the control group. Myocardial revascularization was performed using PCI (92% of patients) or coronary artery bypass grafting (CABG) (8%). Pharmacological support was given on the basis of positive inotropic drugs and vasopressors (remaining unchanged over 30 min after implantation of devices) without further regulations by protocol. Mortality during the first 30 days included 6/13 patients (46%) in the intervention group and 6/13 patients (46%) in the control group (OR 1.00, 95% CI 0.21 to 4.67). During the hospital stay 5/13 patients (38%) in the intervention group and 7/13 patients (54%) in the control group died (OR 0.54, 95% CI 0.11 to 2.55). Six months after randomisation a total of 6/13 patients (46%) in the intervention group and 8/13 patients (62%) in the control group had died (OR 0.54, 95% CI 0.11 to 2.55). The resulting HR of the all-cause 30-day mortality distribution (HR 0.99, 95% CI 0.31 to 3.15) and 6-month mortality distribution (HR 0.73, 95% CI 0.25 to 2.11, log-rank P = 0.55) (Figure 2) reflected no significant differences between groups. Adjustments for age, sex and diabetes showed little influence on these results. During the in-hospital stay no patients with non-fatal cardiovascular events were documented in the intervention group but one patient (4%) with three non-fatal cardiovascular events (reinfarction, reocclusion and subsequent revascularization) was reported in the control group. The primary endpoint of the study was the change of CI 30 minutes after implantation. Pre- implantation CI remained stable in the intervention group (1.73 ± 0.59 to 1.84 ± 0.71 L/min/m^2 post-implantation, P = 0.25) and was noted to rise in the control group (1.71 ± 0.45 to 2.20 ± 0.64 L/min/m^2 post-implantation, P = 0.003). Pre-implantation MAP remained stable in the intervention group (72 ± 17 to 71 ± 22 mmHg post-implantation, P = 0.79) and was noted to rise in the control group (78 ± 16 to 87 ± 18 mmHg post-implantation, P = 0.039). Pre-implantation PCWP tended to decrease in the intervention group (21.9 ± 6.6 to 20.2 ± 5.5 mmHg post-implantation, P = 0.08) as well as in the control group (22.1 ± 8.1 to 19.3 ± 4.7 mmHg post-implantation, P = 0.09). Length of in-hospital stay was 18 ± 11 days in the intervention group and 14 ± 4 days in the control group. There were 3/13 patients with complications (infection in 23% of patients) in the intervention group, and one patient with bleeding and one patient with acute limb ischaemia requiring surgery after device explantation in the control group. No complication could be directly attributed to the use of the devices.

Prondzinsky 2010 between 2003 and 2004 randomised 45 patients in a single-centre open label trial into two groups, a control group (22) and an intervention group (23). Of the intervention group, four patients were excluded (two patients did not fulfil the shock criteria; in one patient the time from MI to shock was ≥ 48 hr; and for one patient no post-randomisation data were available for technical reasons). Among the 22 patients randomised to the control group, one patient was excluded because he did not fulfil the criteria for cardiogenic shock. The intervention group got IABP percutaneously according to standard procedures, via the femoral artery using an 8-French sheath immediately after PCI. Aortic counterpulsation was continued for a minimum of 48 hours. Mean duration of support with IABP was 45 ± 34 hours in the intervention group and 184 hours in the one cross-over patient in the control group. Myocardial revascularization was performed using PCI in 90% (in 85% plus stenting) of patients. Pharmacological support was given on the basis of inotropic and vasopressor agents, aspirin, glycoprotein-IIb or IIIa receptor-blocker, heparin according to standard intensive care guidelines. Mortality during the first 30 days included 6/19 patients (32%) in the intervention group and 5/21 patients in the control group (24%) (OR 1.48, 95% CI 0.37 to 5.96). During the hospital stay 7/19 patients (37%) died in the intervention group and 6/21 patients (29%) in the control group (OR 1.46, 95% CI 0.39 to 5.51). Six months after randomisation a total of 8/17 patients in the in-

tervention group (47%) and 6/18 patients in the control group (33%) had died (OR 1.78, 95% CI 0.45 to 6.97). One year after randomisation a total of nine patients (56%) in the intervention group and six patients (33%) in the control group had died (OR 2.57, 95% CI 0.64 to 10.34). The HRs of the all-cause 30-day mortality distribution (HR 1.55, 95% CI 0.47 to 5.08) and 12-month mortality distribution (HR 1.71, 95% CI 0.63 to 4.63, log-rank P = 0.28) (Figure 2) reflected no significant difference between groups. Adjustments for age, sex and diabetes showed little influence on these results. Eleven patients (28%) with 22 non-fatal cardiovascular events were documented during the in-hospital stay: in the intervention group one patient had reinfarction and recurrent ischaemia and two patients (11%) had reocclusion and subsequent re-revascularization; eight patients (38%) had reocclusion and subsequent re-revascularization in the control group. Mean CI was noted to stay nearly constant in the intervention group with a high variability of post-interventional values (2.32 ± 0.57 to 2.93 ± 1.42 L/min/m^2post-intervention, P = 0.93) as well as in the control group (1.73 ± 0.37 to 2.44 ± 0.67 L/min/m^2 post-intervention, P = 0.23). Mean MAP showed comparable mall changes and high variability in the intervention group (81 ± 11 to 76 ± 16 mmHg post-intervention, P = 0.64) as well as in the control group (83 ± 17 to 80 ± 16 mmHg post-intervention, P = 0.46). Mean PCWP showed no changes but high variability in the intervention group (20.1 ± 5.3 to 20.9 ± 4.4 mmHg post-intervention, P = 0.96), and on a lower level in the control group (14.9 ± 5.6 to 16.2 ± 5.1 mmHg post-intervention, P = 0.78). Length of in-hospital stay was 18 ± 14 days in the intervention group and 29 ± 29 days in the control group. The length of intensive care requirement was 8 ± 7 days in the intervention group and 14 ±12 days in the control group (P = 0.06). There was only one possibly related complication (leg ischaemia) in the intervention group. No patient developed other complications such as bleeding, vascular injury, embolism, infection or thrombocytopenia that could be attributed to IABP use.

Thiele 2012 between 2009 and 2012 randomised 600 patients in a multicentre, open label trial into two groups, a control group (299) and an intervention group (301). Both groups received early revascularization and optimum medical therapy; the intervention group got IABP, the control group no IABP. One patient assigned to IABP was lost to follow-up before 30 days and one patient in the control group withdrew consent. Three additional patients (one from the IABP group and two from the control group) were lost to follow-up before six months. The intervention group got IABP via the femoral artery, and a sheathless insertion was recommended. Mean duration of support was 3.0 days (Interquartile range (IQR) 2.0 to 4.0 days, range 1 to 16 days). Myocardial revascularization was performed using PCI (95.8% of patients) or CABG (3.5% of patients). No revascularization was performed in 3.2% of patients. Pharmacological support was given by haemodynamic monitoring for optimal adjustment of fluid administration and inotropic drugs. All additional treatments were performed according to the standards of the German-Austrian S3-Guidelines (Werdan 2012). During the hospital stay 107/301 patients (36%) in the intervention group and 116/299 patients (39%) in the control group died (OR 0.87, 95% CI 0.62 to 1.21). Mortality during the first 30 days included 119/300 patients (40%) in the intervention group and 123/298 patients (41%) in the control group (OR 0.94, 95% CI 0.67 to 1.30). Six months after randomisation a total of 146/299 patients (49%) in the intervention group and 146/296 patients (49%) in the control group had died (OR 0.98, 95% CI 0.71 to 1.35). One year after randomisation a total of 155/299 patients (52%) in the intervention and 152/296 patients (51%) in the control group had died (OR 1.02, 95% CI 0.74 to 1.41). The resulting HR of the all-cause 30-day mortality distribution (HR 0.93, 95% CI 0.72 to 1.19) and over 12 months (HR 1.01, 95% CI 0.81 to 1.25, log-rank P = 0.94) (Figure 2) reflected no significant difference between groups. Adjustments for age, sex and diabetes showed little influence on these results. In the intervention group, nine patients (3.0%) suffered from reinfarction, four patients (1.3%) had stent thrombosis and two patients (0.7%) suffered from stroke during the in-hospital stay; four patients (1.3%) suffered from reinfarction, three patients (1.0%) had stent thrombosis and five patients (1.7%) suffered from stroke in the control group during the in-hospital stay. Median MAP showed comparable changes and high variability in the intervention group (69 mmHg (IQR 59 to 80) to 73 mmHg (IQR 63 to 87) post-revascularization) and control group (68 mmHg (IQR 59 to 80) to 73 mmHg (IQR 63 to 84) post-revascularization). The length of intensive care unit treatment was 6.0 days (IQR 3 to 13) in both groups. Rates of potential IABP-related complications showed no difference between groups. In the intervention group, 13 patients (4.3%) had peripheral complications requiring intervention, 10 patients (3.3%) suffered from life-threatening or severe bleeding and 52 patients (17.3%) from moderate bleeding. In the control group, 10 patients (3.4%) suffered from peripheral complications requiring intervention and 13 and 49 patients (4.4% and 16.4%) had life-threatening or severe bleeding, or moderate bleeding respectively.

Participants

The age of the patients in the study population of all trials ranged from 28 to 89 years. The proportion of male patients was between 65% and 81%. Between 16% and 54% of participants had diabetes, the percentage of participants with previous infarction was between 22.1% and 58%. The distribution of other baseline characteristics and haemodynamic parameters of patients included in the RCTs are presented in Characteristics of included studies.

Patients were included in eligible RCTs between 1996 and 2012. Between 2 and 274 patients were included per year. Burkhoff 2006 randomised 33 patients with cardiogenic shock due to AMI in 70% of patients, or decompensated chronic heart failure in most of the remaining patients. We restricted our survival analysis to the individual patients with AMI (10 IABP patients and 11 Tandem-

Heart patients). In the study by Ohman 2005 patients with myocardial infarction complicated by hypotension, suspected cardiogenic shock or heart failure were included for randomisation. At the time of randomisation, 22 patients (39%) had Killip class IV. We restricted our analysis to these patients with cardiogenic shock (12 IABP patients and 10 patients in the control group without IABP). In total, 790 patients were included for meta-analysis, of whom 406 patients were treated with IABP and 384 were treated without IABP.

Interventions

Four studies which included 702 patients with AMI and cardiogenic shock compared the intervention IABP versus standard treatment without IABP (Arias 2005; Ohman 2005; Prondzinsky 2010; Thiele 2012) and three studies with 88 randomised patients compared the intervention IABP with percutaneous left ventricular assist devices (LVAD). Two of these studies, with 62 patients, compared IABP versus TandemHeart (Burkhoff 2006; Thiele 2005) and one study with 26 patients compared IABP versus Impella (Seyfarth 2008). The TandemHeart LVAD (TandemHeart, Cardiac Assist, Pittsburgh, PA, USA) is a percutaneous left atrial to femoral arterial LVAD driven by a low-speed centrifugal continuous flow pump (Thiele 2001). The Impella LVAD (Impella LP2.5, Abiomed Europe GmbH, Aachen, Germany) is a catheter-based, impeller-driven, axial flow pump which pumps blood directly from the left ventricle into the ascending aorta (Henriques 2006). In this review the intervention group included all treatment groups with patients randomised to get IABP and the control group included all treatment groups without IABP.
Combining the results of four trials with available data (Ohman 2005; Prondzinsky 2010; Seyfarth 2008; Thiele 2005), the mean duration of support with IABP was 59 ± 57 hours in the IABP group. Myocardial revascularization was performed using PCI in all studies. In total, 95% of patients were revascularized with PCI. Only few patients (5%) underwent thrombolysis as the primary reperfusion strategy, in the trial by Ohman 2005.

Outcomes

Primary data were available to perform a meta-analysis on mortality distribution and mortality rates at discharge from hospital, 30 days, six months and one year after randomisation; haemodynamics; length of hospital and ICU stay; and IABP-related complications. Possible IABP-related complications were reported in different ways and a pooled analysis was only possible for bleeding, vascular injury, leg or limb ischaemia, embolism, infection and thrombocytopenia. Complications in the single trials were described in Included studies. Quality of life was described on the basis of data from one study (Thiele 2012). Time of follow-up varied between time in coronary unit and 10 to 15 days (Arias 2005), 30 days (Burkhoff 2006; Thiele 2005), six months (Ohman 2005; Seyfarth 2008) and one year (Prondzinsky 2010; Thiele 2012).

Excluded studies

Fifteen of the investigated trials did not use a randomised allocation (Anderson 1997; Barron 2001; Bengtson 1992; Gu 2010; Kovack 1997a; Moulopoulos 1986a; Sanborn 2000; Stomel 1994; Stub 2011; Taguchi 2000; Vis 2007a; Vis 2007b; Waksman 1993; Zeymer 2011; Zeymer 2013).
Most RCTs on IABP excluded patients with cardiogenic shock (Christenson 1997a; Christenson 1997b; Christenson 1997c; Christenson 1999; Christenson 2003; Flaherty 1985; Gu 2011; Kono 1996; Ohman 1994; Onorati 2005; Perera 2009; Stone 1997; Vijayalakshmi 2007) or pre-specified cross-over to IABP in the case of cardiogenic shock (Van 't Hof 1999). One trial excluded patients with AMI (O'Neill 2012) and in two trials the patients had no cardiogenic shock at the time of randomisation (Li 2007; Marra 2002). In one trial (O'Rourke 1981) only four patients suffered from cardiogenic shock. Finally, one trial (RECOVER II Trial) included only one patient due to protocol challenges caused by insertion, no cross-over option, consent issues and ethical concerns. The reasons for exclusion are presented briefly in tabular form (please see Characteristics of excluded studies).

Risk of bias in included studies

All trials were published in peer-reviewed journals. Six trials acknowledged the support of either Datascope (Ohman 2005; Prondzinsky 2010), Cardiac Assist (Burkhoff 2006; Thiele 2005), Abiomed Europe GmbH (Seyfarth 2008) or Maquet Cardiopulmonary and Teleflex Medical (Thiele 2012). Datascope, Maquet Cardiopulmonary and Teleflex Medical are manufacturers of the IABP; Cardiac Assist of the TandemHeart LVAD; and Abiomed Europe GmbH developed the Impella LVAD. The range of the number of included participants was 26 to 600. Four trials compared IABP to percutaneous LVAD, three trials to standard treatment without IABP. In five trials the analysis was done by ITT (Burkhoff 2006; Ohman 2005; Prondzinsky 2010; Seyfarth 2008; Thiele 2005; Thiele 2012). Figure 3 and Figure 4 present the risk of bias in the seven eligible studies and summarize risk of bias.

Figure 3. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.

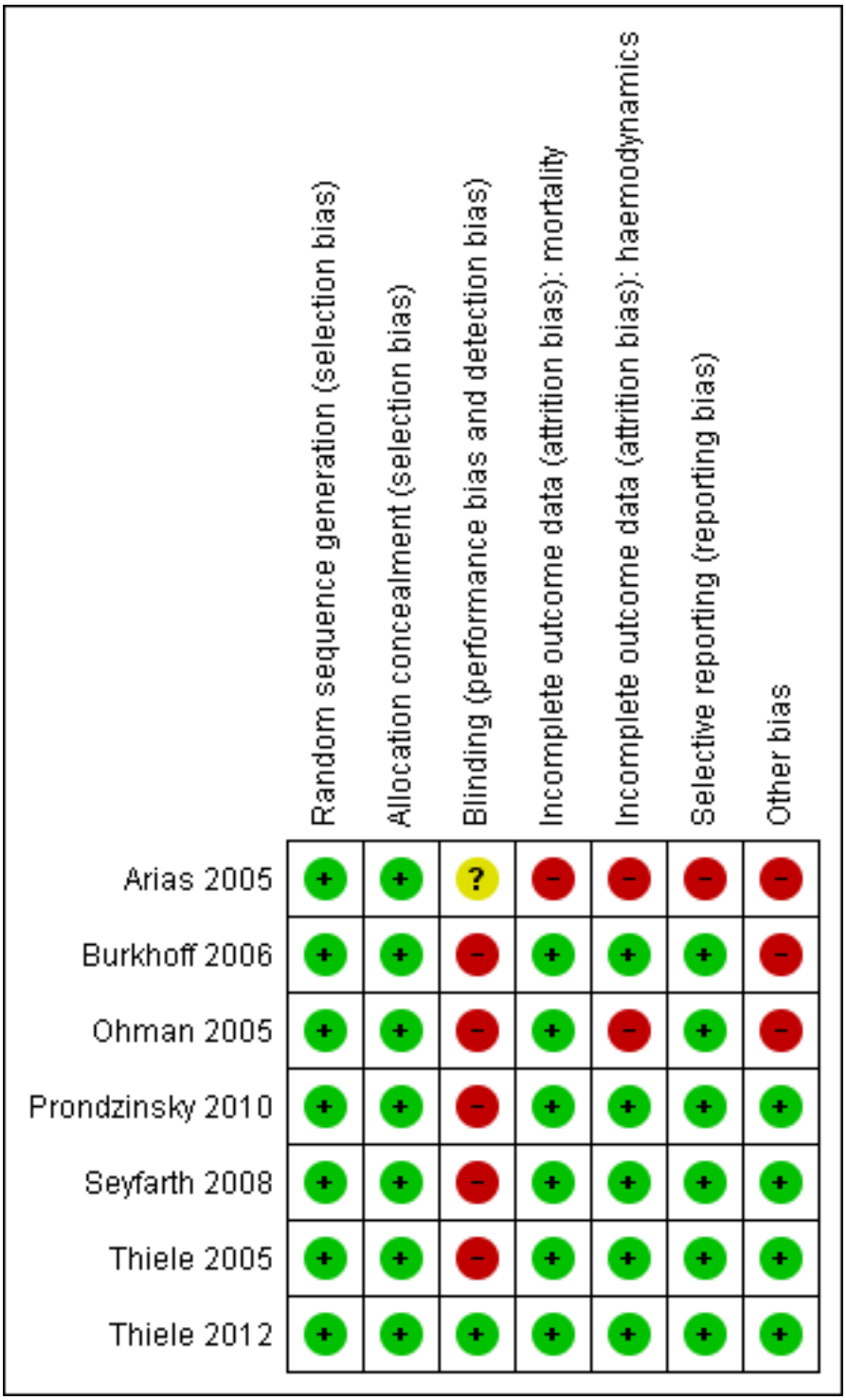
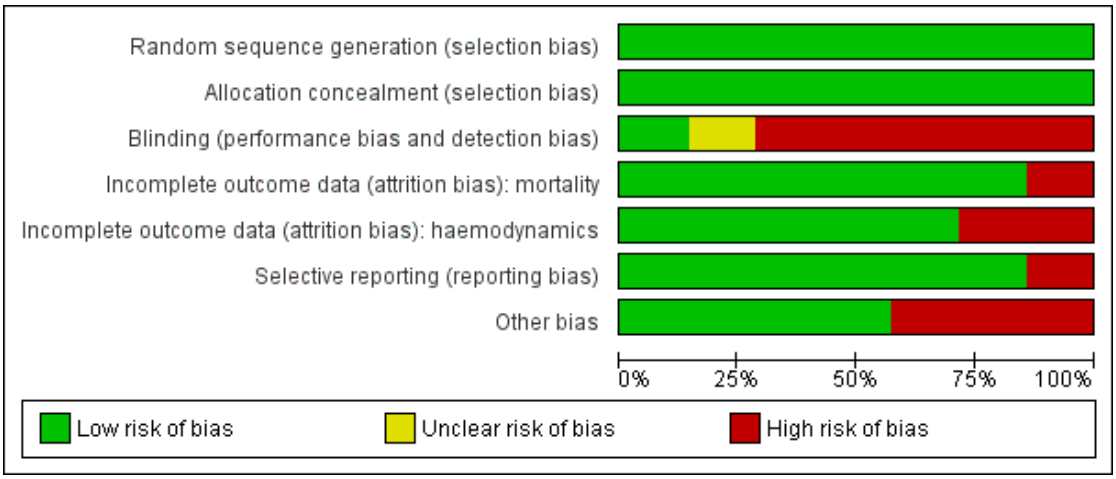


Figure 4. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.



Sequence generation

In all trials currently included in this review the method of sequence generation was provided by the author. All trials used random tables. One trial used random number tables without further restriction (Thiele 2005), two trials used a stratified randomisation (Arias 2005; Thiele 2012) and four trials used a blocked randomisation technique (Burkhoff 2006; Ohman 2005; Prondzinsky 2010; Seyfarth 2008).

Allocation

In all trials currently included in this review the method of allocation concealment was either described in the text or this information was provided by the author. Adequate methods of allocation (opaque sealed envelopes or central telephone allocation) were described in all studies. Five trials used opaque sealed envelopes and in one a central telephone allocation system was used (Ohman 2005).

Blinding

Only one trial described blinding in the study without further detailed information (Arias 2005). Blinding of the intervention to study personnel was not possible introducing the risk of differential behaviour of healthcare providers in all trials. Unblinding of outcome assessment of objective (especially all-cause mortality) outcomes is unlikely to introduce bias.

Incomplete outcome data

All-cause mortality and haemodynamics were investigated.
Arias 2005 restricted reporting of all-cause mortality to in-hospital mortality. Complete 30-day follow-up data were available in five studies. In Thiele 2012 two patients (0.3%) were lost to the 30-day follow-up. Six-month follow-up data for the all-cause mortality distribution were available in four trials (Ohman 2005; Prondzinsky 2010; Seyfarth 2008; Thiele 2012) with missing follow-up data for 10/683 (1.4%) patients. Only two of these studies (Prondzinsky 2010; Thiele 2012) reported 12-month survival status with missing follow-up data for 13/640 (2.0%) patients.
Haemodynamic post-interventional data were reported in four trials (Prondzinsky 2010; Seyfarth 2008; Thiele 2005; Thiele 2012) with follow-up times ranging from post-revascularization (Thiele 2012) to 28 days (Prondzinsky 2010). IPD of three trials (Prondzinsky 2010; Seyfarth 2008; Thiele 2005) were included in the analysis. Follow-up information was lost for 11/106 patients (10.4%, CI), 5/106 (4.7%, MAP) and 16/106 (15.1%, PCWP).

Selective reporting

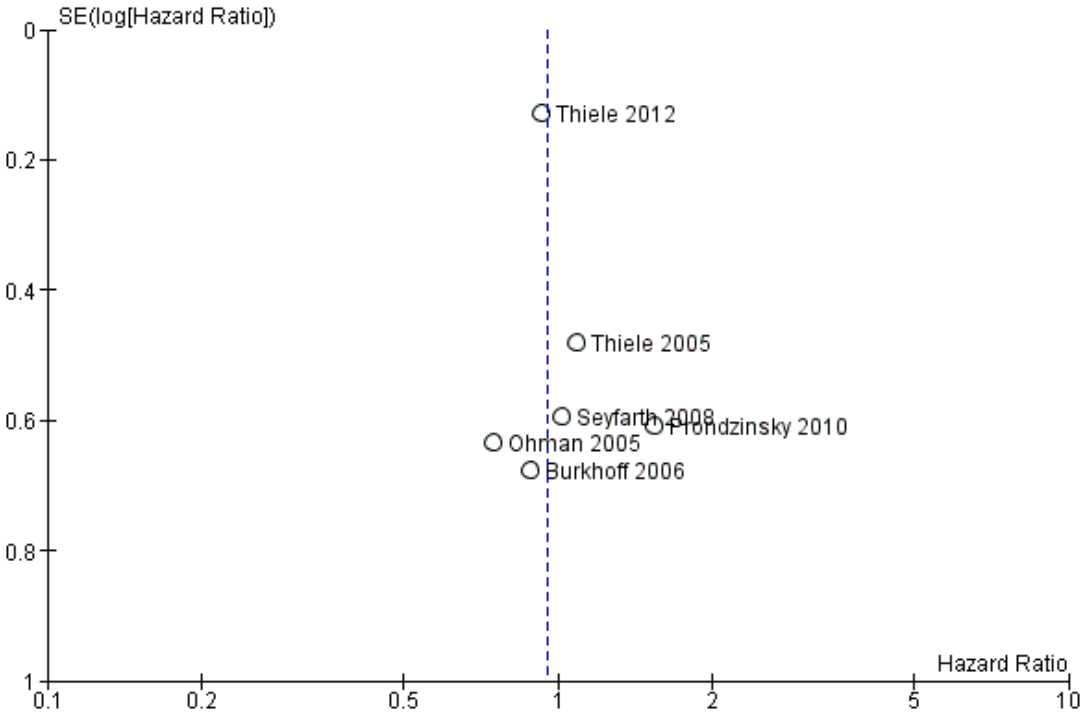
Key outcomes of mortality, haemodynamic parameters and adverse events were reported in six of seven trials. Any information on haemodynamics after randomisation was missed in Arias 2005.

Other potential sources of bias

In three studies (Arias 2005; Burkhoff 2006; Ohman 2005) important deviations from the study plan, which were possible reasons for bias, were documented: high cross-over rates, early stopping, and the inclusion of patients with IABP at randomisation. Cross-over rates were high in these three studies: Arias 2005 reported the results of the per protocol analysis with cross-over of 11/20 patients to the intervention group. Only results of the per protocol analysis were available. These results were restricted to the analysis of in-hospital mortality rates. In Ohman 2005 3/10 patients (33%) from the control group crossed over to the intervention group, and 3/12 (25%) from the intervention group to the control group. The results of the ITT and per protocol (PP) analysis did not show relevant differences. During the first 30 days 6/12 patients (50%) died in the group randomised to IABP and 6/10 patients (60%) in the group randomised to non-IABP (ITT analysis). From the non-survivors, 5/12 patients (42%) died in the group treated with IABP and 7/10 patients (70%) in the group treated without IABP (PP analysis). The resulting ORs were 0.67 (95% CI 0.12 to 3.64) (ITT analysis) and 0.30 (95% CI 0.05 to 1.80) (PP analysis). In Burkhoff 2006 5/14 patients (36%) randomised to IABP and 7/19 patients (37 %) from the control group with LVAD were bridged to another therapy, no patient was bridged to IABP. The overall proportion of cross-over compared with observed event risk was too low to have a clinically relevant impact on the intervention effect estimate for mortality. Most patients (66%) in Burkhoff 2006 were enrolled after failure of IABP before enrolment and randomisation, but all patients still met the haemodynamic criteria for cardiogenic shock. The trial was stopped early on the recommendation of the Data Safety Monitoring Board. Haemodynamic effects were superior in the TandemHeart group compared with the IABP group and it was deemed unlikely to enrol a sufficient number of patients in a reasonable time frame to achieve a more definitive answer concerning mortality. Being aware of these methodological restrictions, we nevertheless decided to include all studies with randomisation of patients with the predefined indication because of the limited number of trials available for this comparison. Risk of bias tables of all single trials are given in detail in Characteristics of included studies.

The funnel plot (Figure 5) was nearly symmetrical for the 30-day mortality distribution with no evidence of publication bias.

Figure 5. Funnel plot of comparison: IABP versus control, outcome: all-cause 30-day mortality distribution.



Effects of interventions

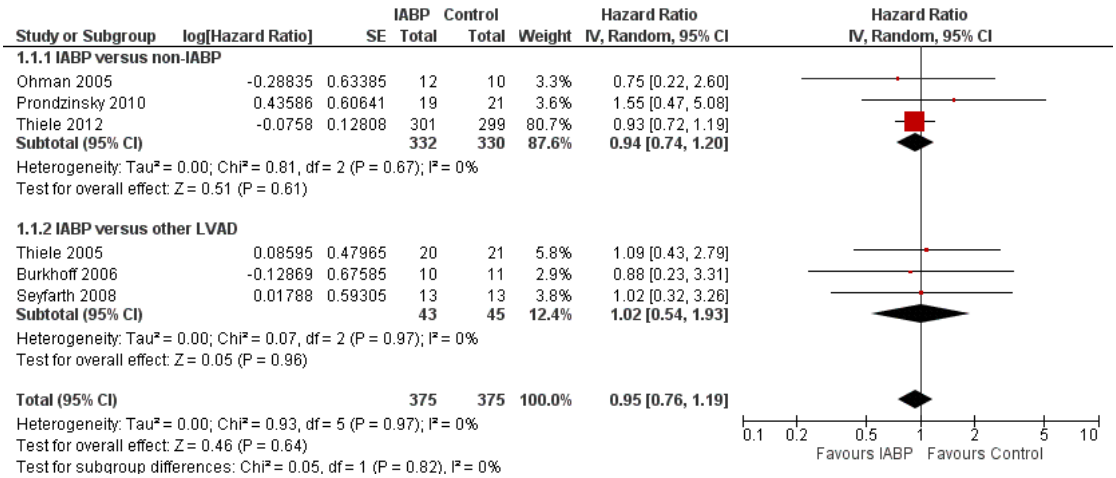
Primary outcome measures

(a) All-cause mortality

Thirty-day all-cause mortality

Across all trials, 6 trials with 750 patients (Burkhoff 2006; Ohman 2005; Prondzinsky 2010; Seyfarth 2008; Thiele 2005; Thiele 2012) reported 150/374 (40.1%) deaths in the intervention group compared with 153/374 (40.9%) deaths in the control group (Analysis 1.4). The HR indicated no difference in mortality distribution between groups (HR 0.95, 95% CI 0.76 to 1.19, Figure 6). This result was consistent whether a one- or a two-step approach was used. Adjustments for age, sex and diabetes did not change the result (Table 1). There was only small heterogeneity observed in these analyses.

Figure 6. Forest plot of comparison: IABP versus control, outcome: all-cause 30-day mortality distribution.



A preplanned analysis according to the types of revascularization also showed no differences between groups. Most patients in six studies (Arias 2005; Burkhoff 2006; Prondzinsky 2010; Seyfarth 2008; Thiele 2005; Thiele 2012) were revascularized by PCI, and in Ohman 2005 patients were revascularized with thrombolytic therapy. An additional sensitivity analysis comparing the influence of the control group intervention (standard without IABP or other LVAD) showed no influence on the results.

A preplanned subgroup analysis investigated the influence of the prognostic factors age and sex on all-cause mortality on the basis of IPD. Combining results across all trials, 97/252 (38.5%) men and 53/123 (43.1%) women died in the intervention group and 107/264 (40.5%) men and 49/111 women (44.1%) in the control group. In total, 86/251 (34.3%) of patients < 75 years and 64/124 (51.6%) of patients ≥ 75 years died in the intervention group, and 108/281 (38.4%) and 48/94 (51.1%) in the control group, respectively (Table 1).

In-hospital all-cause mortality

The meta-analysis was conducted on 5 trials with 747 patients to describe in-hospital mortality rates (Arias 2005; Prondzinsky 2010; Seyfarth 2008; Thiele 2005; Thiele 2012). Across trials, 139/384 patients (36.2%) died in the intervention group and 143/363 patients (39.4%) in the control group (OR 0.87, 95% CI 0.65 to 1.18) (Analysis 1.3) with no heterogeneity between trials.

Six-month and 12-month all-cause mortality

The meta-analysis on six-month mortality was conducted on four trials with 678 patients to describe all-cause six month mortality (Ohman 2005; Prondzinsky 2010; Seyfarth 2008; Thiele 2012). Across trials, 166/341 patients (48.7%) died in the intervention group and 167/337 patients (49.6%) in the control group (Analysis 1.5).

One-year mortality information was available from 627 patients of two trials (Prondzinsky 2010; Thiele 2012). In total, 164/315 patients (52.1%) died in the intervention group and 158/312

patients (50.6%) in the control group.
The HR of 1.02 (95% CI 0.84 to 1.25) indicated no difference in mortality distribution between groups, even when adjustments were made for age, sex and diabetes; or a one- or two-step approach was used (Table 1). There was no substantial heterogeneity observed in these analyses.
A preplanned analysis according to the types of revascularization also showed no differences between groups. Most patients in Prondzinsky 2010, Seyfarth 2008 and Thiele 2012 were revascularized by PCI, and in Ohman 2005 patients were revascularized with thrombolytic therapy. An additional sensitivity analysis comparing standard care without IABP or other LVAD did not influence the results.

(b) Non-fatal cardiovascular events

Data were available in 5 trials with 727 patients (Ohman 2005; Prondzinsky 2010; Seyfarth 2008; Thiele 2005; Thiele 2012). Reported results were combined across trials. During hospitalisation, 11/364 patients (3.0%) from the intervention group and 5/363 patient (1.4%) from the control group suffered from reinfarction; 4/364 patients (1.1%) from the intervention group and 5/363 (1.4%) patients from the control group had a stroke. Reocclusion and subsequent re-revascularization were reported in 4 trials with 705 patients (Prondzinsky 2010; Seyfarth 2008; Thiele 2005; Thiele 2012). Altogether, 6/352 patients (1.7%) from the intervention group and 13/353 patients (3.7%) from the control group had re-occlusion with subsequent re-revascularization. Recurrent ischaemia was reported in 2 trials with 81 patients (Prondzinsky 2010; Thiele 2005). Overall, 2/39 patients (5.1%) from the intervention group and 0/42 patients from the control group had recurrent ischaemia (Table 2).

Secondary outcome measures

(a) Haemodynamics (cardiac index, mean arterial pressure, pulmonary capillary wedge)

Cardiac index (CI)

The CI measured after device implantation was available for 95 patients from 3 eligible studies (Prondzinsky 2010; Seyfarth 2008; Thiele 2005), with substantial heterogeneity (I^2 = 85%) between trials. To explore the heterogeneity, a subgroup analysis according to the comparison group was conducted and showed relevant differences according to comparison group (non-IABP versus other LVAD) on the results. In Prondzinsky 2010, patients randomised to IABP (n = 16) had a higher mean CI compared to control group patients (n = 14) without any assist devices (mean difference (MD) 0.49 L/min/m^2, 95% CI -0.29 to 1.27). In contrast, combining the results of 2 trials, patients randomised to IABP (n = 32) had lower mean CI compared to control group patients (n = 32) with LVAD (MD -0.75 L/min/m^2, 95% CI -1.51 to 0.00).

Mean arterial pressure (MAP)

The MAP after device implantation was available on 101 patients from 3 eligible studies. Combining across trials, patients in the intervention group (n = 50) showed lower mean MAP values post-implantation compared to control group patients (n = 51) with high variation between patients (MD -5.1 mmHg, 95% CI -10.9 to 0.66) and low heterogeneity between trials.
These differences were not stated in Thiele 2012. Median MAP showed comparable changes and high variability in both groups. In the intervention group, MAP increased from median 69 mmHg (IQR 59 to 80) to 73 mmHg (IQR 63 to 87) post-revascularization, and patients in the control group showed comparable changes from median 68 mmHg (IQR 59 to 80) to 73mmHg (IQR 63 to 84) post-revascularization.

Pulmonary capillary wedge pressure (PCWP)

The PCWP after device implantation was available for 90 patients from 3 eligible studies. Combining across trials, patients in the intervention group (n = 45) showed higher mean PCWP values post-implantation compared to control group patients (n = 45) (MD 3.9 mmHg, 95% CI 1.1 to 6.7), with moderate heterogeneity between trials (I^2 = 44%). To explore the heterogeneity a subgroup analysis according to the comparison group (non-IABP or other LVAD) was conducted but a difference according to comparison group was not found.

(b) Length of hospital and intensive care unit (ICU) stay

Information about length of hospital stay was available from 4 trials with 677 patients (Prondzinsky 2010; Seyfarth 2008; Thiele 2005; Thiele 2012) and information about ICU requirements was available from 640 patients in 2 trials (Prondzinsky 2010; Thiele 2012). Combining trials, no difference in length of in-hospital stay was shown between the two treatment groups (MD -1.1 days, 95% CI -4.9 to 2.7) with low heterogeneity between trials (I^2 = 24%). Prondzinsky 2010 showed a benefit in the ICU requirement with a shorter mean time in the intervention group (n = 19) compared to patients in the control group (n = 21) (MD -6.2 days, 95% CI -12.3 to -0.07), but this difference was not seen by Thiele 2012 with a median of 6 days of ICU treatment in both groups.

(c) Quality of life

Information on quality of life was available on 1-year survivors of one trial with 600 patients (Thiele 2012). They reported information on health-related quality of life and assessed symptoms of heart failure according to the New York Heart Association (NYHA) and angina according to the Canadian Cardiovascu-

lar society (CCS) classification. Health-related quality of life was described by the EQ-5D-3L index value in 5 dimensions (mobility, self-care, usual activities, pain or discomfort, and anxiety or depression) and was assessed for 274/286 (95%) of the 1-year survivors. Less than 10% of patients described severe problems, especially in usual activities, self-care and anxiety or depression, with no differences between treatment groups. NYHA and CCS class were recorded in 253 and 252 (both 88%) of the 286 1-year survivors. Of these patients 115/127 (91%) from the intervention group and 118/126 (94%) from the control group were in NYHA class I or II. In total, 125/127 (98%) in the intervention group and 124/125 (99%) in the control group were in CCS class I or II.

(d) All IABP-related post-interventional complications

Possible IABP-related complications were described heterogeneously in the trials. Analyses displayed frequencies of possibly related complication such as moderate or severe bleeding, vascular injury, leg or limb ischaemia, embolism, infection and thrombocytopenia. In detail, a high incidence of complications in the control groups had to be attributed to interventions with other LVAD. As a consequence, the frequency in the intervention group versus control groups with LVAD and without LVAD were analysed separately (Table 3). Reported results were combined across trials. Six trials (Burkhoff 2006; Ohman 2005; Prondzinsky 2010; Seyfarth 2008; Thiele 2005; Thiele 2012) reported the frequency of post-interventional complications.

In total, 27/378 patients (7.1%) from the intervention group and 21/329 (6.4%) patients in the control group without LVAD, but 27/53 patients (51%) from the control group with LVAD, suffered from moderate or severe bleeding. None of the 33 patients in the intervention group and 21 patients in the control group without LVAD, but 2/19 patients (10%) from the control group with LVAD, suffered from vascular injury. Three of 78 patients (3.8%) from the intervention and 1/31 patient (3.2%) from the control group without LVAD, but 12/53 patients (23%) from control groups with LVAD, had leg or limb ischaemia. Embolism was reported in 2 trials, 1/40 patient (2.5%) from the intervention group but no patient from control groups suffered from embolism. Fifteen of 66 patients (23%) from the intervention group and none of 21 patients from the control group without LVAD, but 17/53 patients (32%) from the control group with LVAD, suffered from infection. Thrombocytopenia was reported in three trials. Three of 53 patients (5.7%) from the intervention group and none of 21 patients from the control group without LVAD, but 4/40 patients (10%) from the control group with LVAD, suffered from thrombocytopenia. Peripheral ischaemic complications requiring intervention were reported in 1 trial, where 12/300 (4.3%) from the intervention group and 10/298 (3.4%) patients from the control group without LVAD suffered from a peripheral ischaemic complication.

DISCUSSION

Summary of main results

Data from seven eligible studies with a total of 790 patients with AMI and cardiogenic shock that compared IABP versus no IABP could not show convincing evidence for either benefit or harm to support the use of intra-aortic balloon pump counterpulsation (IABP). Six of these analysed studies were too small to be sufficiently powered to investigate the beneficial or harmful effects of IABP support beyond initial haemodynamic improvements. One trial (Thiele 2012) was powered to discover a 12% absolute difference in the 30-day survival rates reported in registries and meta-analyses. This study and even the aggregated study population of all included trials did not show a significant reduction of mortality in patients with AMI and cardiogenic shock.

Four trials compared IABP to standard treatment and three to other percutaneous LVAD. Combining the result of all six trials, in the intervention and control arms 40% and 41% of patients died during the first 30 days and 49% and 50% of patients died during the 6 months after randomisation, respectively. Hazard ratios (HRs) of the all-cause mortality distribution provide no evidence for a survival benefit. Adjustments for age, sex and diabetes in subgroups (IABP versus non-IABP and IABP versus other LVAD) confirmed these results. While differences in survival were comparable in patients treated with IABP, compared to with and without LVAD, haemodynamic results from three small studies showed heterogeneous results. Post-implantation, patients randomised to IABP had a higher mean CI than patients without assist devices, and lower mean CI than patients with LVAD. Patients in the IABP group showed lower MAP values and higher mean PCWP. Differences in MAP were not stated in the larger trial, whereas CI and PCWP were not reported.

A higher incidence of complications was observed in control groups with other LVADs, especially in the frequency of moderate and severe bleeding, compared to the intervention group with IABP. Results on quality of life are in line with former published follow-up data of the SHOCK Trial (Sleeper 2005), where survivors of infarct-related cardiogenic shock showed a surprisingly good quality of life. Although one-year mortality after emergency revascularization was still high (54%), most survivors had good functional status.

Quality of the evidence

The results of this meta-analysis are limited by several issues concerning the number of trials, the number of patients in most trials, the heterogeneity of included patients at baseline, and the fact that IABP has been compared to no-IABP and to other LVADs.

a) While the patient numbers in previous results have been too small to address mortality, this review now includes individual patient data (IPD) of the latest and also largest RCT on this topic

(Thiele 2012). For this reason, this updated review comprises the latest available evidence from RCTs to evaluate the therapeutic effects on mortality of IABP support in infarct-related cardiogenic shock. Additionally, it can be looked upon as a positive aspect that all included trials have been performed by different investigators at different institutions, so that bias due to repeated single-centre experiences can be excluded.

b) All these analysed trials have been conducted properly under the conditions of prospective RCTs, so that this meta-analysis can contribute valuable data to estimate the effects of the IABP support in infarct-related cardiogenic shock, although six of these seven trials were not powered to detect differences in mortality. The evidence which could have been derived by former meta-analyses had been hampered by small patient numbers being included in the trials. This methodological gap now can be closed by the IABP-SHOCK II Trial, which was adequately powered to detect mortality differences under IABP support. More specifically, there are points of concern within most of the included and analysed studies.

There was frequent cross-over in several trials.

Ohman 2005: in total, 33% of patients from the control group with cardiogenic shock crossed over to emergency IABP and 25% of the intervention group with cardiogenic shock did not receive IABP.

Thiele 2005: one patient with rapid haemodynamic improvement did not receive the assigned LVAD.

Arias 2005: altogether 27.5% of patients assigned to the control group crossed over to IABP and were analysed in this group.

Burkhoff 2006: this multicentre trial with a small sample size stopped recruitment earlier than initially intended. Futhermore, different therapeutic strategies were performed in the included patients. Moreover, there was a complex cross-over situation, 37% of patients from the control group were bridged to another therapy (3 patients to LVAD, 1 patient to extracorporeal membrane oxygenation, 2 patients to PCI with stent placement, 1 patient to mitral valve repair).

Prondzinsky 2010: one patient in the control group was treated with IABP.

Seyfarth 2008: one patient assigned to the control group died before implantation and did not receive the LVAD.

c) With the limited number of included trials, there was one trial (Ohman 2005) which included patients with acute decompensated heart failure and haemodynamically instable patients at the same time. In this case only IPD eligible according to the criteria of cardiogenic shock could be extracted and evaluated.

d) Another limitation is given by the fact that only four RCTs compared IABP support versus a control without any device giving haemodynamic support. Three trials compared the IABP support to other assist devices such as the Tandem-Heart or the Impella system.

e) For the limited number of four trials comparing IABP to no IABP, IPD of only three of the four trials were available. One trial included patients being revascularized by thrombolysis and the others by PCI. Indeed, in this meta-analysis there were only 640 patients revascularized by the present state of the art, by primary PCI, and compared to a no IABP control group.

f) Although there has been a favourable trend in haemodynamics for IABP support compared to standard treatment without IABP, it could be shown that systemic inflammation and also multi-organ failure seem to have a higher impact on prognosis in infarct-related cardiogenic shock than haemodynamic parameters (Lim 2003; Prondzinsky 2010). Based on this background the issue of IABP timing becomes of scientific interest.

While the effect of IABP timing in the field of cardiac surgery has been discussed during the last few years (Ramnarine 2005) this discussion currently has been continued in the field of cardiogenic shock (Abdel-Wahab 2010; Cheng 2013). As shown in Lim 2003 the haemodynamic parameters in critically ill cardiogenic shock patients are less predictive than expected. These surprising findings might be explained in particular by the effect of timing (early or late initiation of IABP support). Haemodynamic stabilization under the terms of cardiogenic shock has to be achieved as soon as possible, following the principle of early goal directed therapy, to prevent the prognostically relevant multi-organ dysfunction syndrome (MODS) or multi-organ failure (MOF).

g) Another methodological limitation is the heterogeneity of the investigated patient groups themselves. While patients with acute NSTEMI or STEMI can be described very well regarding their baseline conditions, haemodynamically unstable patients with infarct-related cardiogenic shock show a broader variation of physiological parameters at baseline. Especially so are the markers of MOF and systemic inflammation (SIRS), which have had the greatest impact on prognosis in infarct-related cardiogenic shock (Prondzinsky 2010), and they are often not measured at baseline.

h) Here we refer to section b), where major concerns regarding the issue of cross-over designs were explained.

i) The duration of IABP support differed in the intervention groups in the trials from 26 ± 19 up to 84 ± 54 hours. The different duration of IABP support may reflect different patient populations regarding the degree of haemodynamic instability but may also indicate treatment algorithms for management for cardiogenic shock, regarding inotropic and vasopressor support as well as IABP weaning.

j) As shown by the OASIS-5 trial (Yusuf 2006), the outcome in haemodynamically stable patients with acute coronary syndromes is obviously driven by the (intervention-related) bleeding rate. Therefore it would have been very helpful if the assessment of complications, in particular bleeding, were performed in a comparable way in all analysed trials i.e. by the TIMI-bleeding definition. Therefore, the bleeding-related impact on outcomes could not be determined. It cannot be excluded that bleeding counteracts the trend of favourable haemodynamics in the intervention group.

k) Abdel-Wahab 2010 performed a retrospective analysis of 48 patients with cardiogenic shock and found more favourable results for those cardiogenic shock patients in whom the IABP support

had been initiated prior to primary PCI. Indeed, the issue of timing of IABP support in primary PCI has not been investigated in detail. Therefore, it can not be excluded that an earlier initiation of IABP support might have an impact on outcome, due to earlier increased macro-circulation with improved consecutive microcirculatory disturbances preventing multi-organ dysfunction or failure. However, other trials did not show a benefit with earlier IABP timing (Cheng 2013).

For many years there has been a strong recommendation by the ACC/AHA and also ESC for the use of intra-aortic counterpulsation under the conditions of cardiogenic shock. Instead of these strong recommendations the utilization rate of adjunctive IABP support in STEMI complicated by cardiogenic shock remained low (20% to 39%). This gap of a strong recommendation, predominantly based on non-randomised trials and registries, on one hand and restricted guideline adherence in daily clinical practice on the other hand should now be resolved in the light of the revised guidelines of the ACC/AHA, ESC and also the corresponding German-Austrian Guideline (Steg 2012; O'Gara 2013; Werdan 2012). The impact of the IABP-SHOCK II Trial and this current meta-analysis on guideline recommendations needs to be elucidated in the future.

Additionally, our findings showed an increased rate of bleeding with the use of intra-aortic counterpulsation. Mortality in acute coronary syndrome (ACS) is predominantly driven by bleeding, especially major bleeding, showing a strong relationship to poor outcome in ACS patients. Based on this background the findings of our review might explain why the majority of clinicians, being afraid of significant bleeding, avoided IABP support in cardiogenic shock.

Agreements and disagreements with other studies or reviews

Other randomised studies and reviews

During the last decades several prospective RCTs (Flaherty 1985; Kono 1996; O'Rourke 1981; Stone 1997; Van 't Hof 1999) with IABP support in ST-segment elevation myocardial infarction (STEMI) without cardiogenic shock have been performed. These trials have recently been investigated and analysed in different meta-analyses (Cassese 2012; Cheng 2009; Sjauw 2009). The first published systematic review (Sjauw 2009) included seven RCTs with 1009 patients after STEMI with and without cardiogenic shock. IABP neither showed a 30-day survival benefit nor improved left ventricular function, while IABP support was associated with higher stroke and bleeding rates. Another systematic review compared the results of RCTs comparing percutaneous LVAD with IABP, on the basis of three available trials on patients with cardiogenic shock (Cheng 2009) and stated that 'although use of percutaneous LVAD resulted in a better haemodynamic profile compared with IABP counterpulsation, this did not translate into improved 30-day survival', but patients treated with LVAD tended to have a higher incidence of leg ischaemia and device-related bleeding. This result was stated by Cassese 2012 looking at 1054 patients with AMI without CS from six randomised trials. The present review adds evidence from three additional studies (Arias 2005; Prondzinsky 2010; Thiele 2012) and subgroups of patients with myocardial infarction and cardiogenic shock from two other studies (Burkhoff 2006; Ohman 2005), and represents a formal assessment of mortality distribution.

Other non-randomised studies and reviews

Recent reviews on randomised and non-randomised trials (Bahekar 2012; Romeo 2013; Sjauw 2009; Zhang 2013) came to different conclusions.

Sjauw 2009 conducted a second meta-analysis in 9 cohort studies with 10,529 patients with STEMI and cardiogenic shock. In this meta-analysis, the subgroup treated with thrombolysis showed an 18% (95% CI 16 to 20) decrease in 30-day mortality with IABP support. These findings are limited by higher revascularization rates compared to patients without IABP support. As shown by Hochman 1999, revascularization of the infarct-related artery in infarct-related cardiogenic shock had a relevant impact on outcomes. Additionally, there was a bias towards younger age in the IABP group. For this reason the reported beneficial effects of IABP support in AMI patients after thrombolysis have to be interpreted carefully. On the other hand, in patients treated with PCI, IABP was associated with an increased mortality rate of 6% (95% CI 3 to 10) with IABP support.

Bahekar 2012 conducted a meta-analysis in 6 cohort studies with 24,541 patients with AMI complicated by cardiogenic shock and stated a significant reduction in mortality, by 28% (95% CI 14 to 40), with IABP. Another systematic review (Zhang 2013) analysed 13 RCTs with 1958 patients following AMI. The authors summarized that IABP therapy is effective in reducing earlier mortality post-AMI, particularly for patients with cardiac shock, and reduced 30-day mortality by 35% (95% CI 3 to 56). Unfortunately only the abstract is published in English, while the paper has been published in Chinese so that a broad discussion in the scientific community may be hampered.

In contrast to this result, Romeo 2013 concluded on the basis of 17 cohort studies involving 7407 patients with AMI and cardiogenic shock that IABP support is only effective in patients with thrombolytic therapy (reduction by 23%, 95% CI 13 to 32), but is associated with a significant increase in in-hospital mortality, by 18% (95% CI 4 to 34), in patients revascularized by primary PCI. This review included three RCTs with randomisation of IABP (Ohman 2005; Prondzinsky 2010; Thiele 2012) and 14 observational studies.

In contrast to these beneficial findings, data from the National

Registry for Acute Myocardial Infarction (Spencer 2001), the Euro Heart Survey on PCI (Zeymer 2011) and the German ALKK-PCI registry (Zeymer 2013) showed no hint of beneficial effects with IABP support in patients with infarct-related cardiogenic shock who were revascularized by PCI.

AUTHORS' CONCLUSIONS

Implications for practice

In contrast to the previous version, this review includes data from 790 patients and allows more definite conclusions on the potential beneficial or harmful clinical effects of IABP support beyond the initial haemodynamic improvements. IABP support should no longer be recommended by the guidelines in every case of infarct-related cardiogenic shock, and should be supported as an individual treatment option based on the personal experience and decision of the investigator and the particular circumstances of the individual treatment situation (Werdan 2012).

Implications for research

According to the recently reported data (Prondzinsky 2010) systemic inflammation substantially contributes to the outcome in cardiogenic shock. Though it is common that the first phase of cardiogenic shock is accompanied by compensatory vasoconstriction, recent studies have shown that during the following phases of cardiogenic shock inappropriate vasodilation induced by inflammation seems to be the key for understanding the persisting haemodynamic instability as reflected by increasing rates of use of inotropes and vasopressors (Debrunner 2008; Geppert 2002; Geppert 2006; Hochman 2003; Kohsaka 2006; Seely 2000).

As a consequence, subgroups of patients and the phases of cardiogenic shock regarding systemic inflammation and multi-organ failure clearly have to be defined to allow a better discrimination of patient groups, regardless of which assist device will be investigated. Only the consequent quantifiable evaluation of inflammation and organ failure will allow a reliable interpretation of further trials, to detect beneficial or harmful effects on outcomes in different subgroups.

Future trials investigating haemodynamic support by other active left ventricular assist devices (LVAD) will have to examine whether significant haemodynamic improvements without increased bleeding rates can be provided.

ACKNOWLEDGEMENTS

The excellent support from the Cochrane Heart Group and the German Cochrane Centre is much appreciated.

We would like to acknowledge Daniel Burgdorff, Magnus Ohman and Beth Fraulo for collaboration in the collection of individual patient data and Dr Eduardo Arias who provided unpublished information. We thank Andreas Wiencke, Katharina Hirsch, Alexander Solms for support in statistical analysis regarding frailty models and Doris Gerlach for the support as librarian.

REFERENCES

References to studies included in this review

Arias 2005 *{published and unpublished data}*

Arias EA, González-Chon O, García-López SMC, Chacón MA, Noriega-Iriondo F, Vega RE, et al. Impact of the intra-aortic balloon pump in the mortality due to cardiogenic shock secondary to acute myocardial infarction [Spanish] [Impacto delbalón intra–aórtico de contrapulsación en la mortalidad por choque cardiogénico secundario a infarto agudo delmiocardio]. *Archivos de Cardiologia de Mexico* 2005;**75**(3):260–6.

Burkhoff 2006 *{published and unpublished data}*

Burkhoff D, Cohen H, Brunckhorst C, O'Neill WW. A randomized multicenter clinical study to evaluate the safety and efficacy of the TandemHeart percutaneous ventricular assist device versus conventional therapy with intraaortic balloon pumping for treatment of cardiogenic shock. *American Heart Journal* 2006;**152**(3):469.e1–8. [DOI: 10.1016/j.ahj.2006.05.031]

Ohman 2005 *{published and unpublished data}*

* Ohman EM, Nanas J, Stomel RJ, Leesar MA, Nielsen DWT, O'Dea D, et al. Thrombolysis and counterpulsation to improve survival in myocardial infarction complicated by hypotension and suspected cardiogenic shock or heart failure: results of the TACTICS Trial. *Journal of Thrombosis and Thrombolysis* 2005;**19**(1):33–9.

Talley JD, Ohman EM, Mark DB, George BS, Leinberger JD, et al. Economic implications of the prophylactic use of intraaortic balloon counterpulsation in the setting of acute myocardial infarction. *American Journal of Cardiology* 1997; **79**:590–4.

Prondzinsky 2010 *{published and unpublished data}*

Prondzinsky R, Lemm H, Swyter M, Wegener N, Unverzagt S, Carter JM, et al. Intra-aortic balloon counterpulsation in patients with acute myocardial infarction complicated by cardiogenic shock: The prospective, randomized IABP SHOCK Trial for attenuation of multiorgan dysfunction syndrome. *Critical Care Medicine* 2010;**38**(1):1–9. [DOI: 10.1097/CCM.0b013e3181b78671]

Seyfarth 2008 *{published and unpublished data}*
Seyfarth M, Sibbing D, Bauer I, Fröhlich G, Bott-Flügel L, Byrne R, et al. A randomized clinical trial to evaluate the safety and efficacy of a percutaneous left ventricular assist device versus intra-aortic balloon pumping for treatment of cardiogenic shock caused by myocardial infarction. *Journal of the American College of Cardiology* 2008;**52**(19):1584–8. [DOI: 10.1016/j.jacc.2008.05.065]

Thiele 2005 *{published and unpublished data}*
Thiele H, Sick P, Boudriot E, Diederich K-W, Hambrecht R, Niebauer J, et al. Randomized comparison of intra-aortic balloon support with a percutaneous left ventricular assist device in patients with revascularized acute myocardial infarction complicated by cardiogenic shock. *European Heart Journal* 2005;**26**(13):1276–83. [DOI: 10.1093/eurheartj/ehi161]

Thiele 2012 *{published and unpublished data}*
Thiele H, Schuler G, Neumann F-J, Hausleiter J, Olbrich H-G, Schwarz B, et al. Intraaortic balloon counterpulsation in acute myocardial infarction complicated by cardiogenic shock: Design and rationale of the Intraaortic BalloonPump in Cardiogenic Shock II (IABP-SHOCK II) trial. *American Heart Journal* 163;**6**:938–45. [DOI: 10.1016/j.ahj.2012.03.012]
Thiele H, Zeymer U, Neumann F-J, Ferenc M, Olbrich H-G, Hausleiter J, et al. Intra-aortic balloon counterpulsation in acute myocardial infarction complicated by cardiogenic shock (IABP-SHOCK II): final 12 month results of a randomised, open-label trial. *Lancet* 2013;**382**(9905):1638–45.
* Thiele H, Zeymer U, Neumann F-J, Ferenc M, Olbricht H-G, Hausleiter J, et al. Intraaortic balloon support for myocardial infarction with cardiogenic shock. *New England Journal of Medicine* 2012;**367**(14):1287–96. [DOI: 10.1056/NEJMoa1208410]

References to studies excluded from this review

Anderson 1997 *{published data only}*
Anderson RD, Ohman EM, Holmes DR, Col I, Stebbins AL, Bates ER, et al. Use of intraaortic balloon counterpulsation in patients presenting with cardiogenic shock: observations from the GUSTO-I Study. Global Utilization of Streptokinase and TPA for Occluded Coronary Arteries. *Journal of the American College of Cardiology* 1997;**30**(3):708–15.

Barron 2001 *{published data only}*
Barron HV, Every NR, Parsons LS, Angeja B, Goldberg RJ, Gore JM, et al. The use of intra-aortic balloon counterpulsation in patients with cardiogenic shock complicating acute myocardial infarction: data from the National Registry of Myocardial Infarction 2. *American Heart Journal* 2001;**141**:933-9.

Bengtson 1992 *{published data only}*
Bengtson JR, Kaplan AJ, Pieper KS, Wildermann NM, Mark DB, Pryor DB, et al. Prognosis in cardiogenic shock after acute myocardial infarction in the interventional era. *Journal of the American College of Cardiology* 1992;**20**:1482-9.

Christenson 1997a *{published data only}*
Christenson JT, Simonet F, Badel P, Schmuziger M. Evaluation of preoperative intra-aortic balloon pump support in high risk coronary patients. *European Journal of Cardio-Thoracic Surgery* 1997;**11**(6):1097–103.

Christenson 1997b *{published data only}*
Christenson JT, Simonet F, Badel P, Schmuziger M. The effect of preoperative intra-aortic balloon pump support in patients with coronary artery disease, poor left-ventricular function (LVEF < 40%), and hypertensive LV hypertrophy. *The Thoracic and Cardiovascular Surgeon* 1997;**45**:60–4.

Christenson 1997c *{published data only}*
Christenson JT, Badel P, Simonet F, Schmuziger M. Preoperative intraaortic balloon pump enhances cardiac performance and improves the outcome of redo CABG. *Annals of Thoracic Surgery* 1997;**64**:1237–44.

Christenson 1999 *{published data only}*
Christenson JT, Simonet F, Badel P, Schmuziger M. Optimal timing of preoperative intraaortic balloon pump support in high-risk coronary patients. *Annals of Thoracic Surgery* 1999;**68**:934–8.

Christenson 2003 *{published data only}*
Christenson JT, Licker M, Kalangos A. The role of intra-aortic counterpulsation in high-risk OPCAB surgery: A prospective randomized study. *Journal of Cardiac Surgery* 2003;**18**(4):286–94.

Flaherty 1985 *{published data only}*
Flaherty JT, Becker LC, Weiss JL, Brinker JA, Bulkley BH, Gerstenblith G, et al. Results of a randomized prospective trial of intraaortic balloon counterpulsation and intravenous nitroglycerin in patients with acute myocardial infarction. *Journal of the American College of Cardiology* 1985;**6**(2):434–46.

Gu 2010 *{published data only}*
Gu J, Hu W, Xiao H, Feng X, Chen Y, Zhang D. Intra-aortic balloon pump improves clinical prognosis and attenuates C-reactive protein level in acute STEMI complicated by cardiogenic shock. *Cardiology* 2010;**117**:75–80.

Gu 2011 *{published data only}*
Gu J, Hu W, Xiao H, Feng X, Zhiping S, et al. Prophylactic intra-aortic balloon pump reduces C-reactive protein levels and early mortality in high-risk patients undergoing percutaneous coronary intervention. *Acta Cardiologica* 2011;**66**:499–504.

Kono 1996 *{published data only}*
Kono T, Morita H, Nishina T, Fujita M, Onaka H, Hirota Y, et al. Aortic counterpulsation may improve late patency of the occluded coronary artery in patients with early failure of thrombolytic therapy. *Journal of the American College of Cardiology* 1996;**28**(4):876–81.

Kovack 1997a *{published data only}*
Kovack PJ, Rasak MA, Bates ER, Ohman EM, Stomel RJ. Thrombolysis plus aortic counterpulsation: improved

survival in patients who present to community hospitals with cardiogenic shock. *Journal of the American College of Cardiology* 1997;**29**(7):1454-8.

Li 2007 *{published data only}*
Li JL, Xue H, Wang BS, Zhang HY, Yin LH, Shun WD, et al. Effect of prolonged intra-aortic balloon pumping in patients with cardiogenic shock following acute myocardial infarction. *Medical Science Monitor* 2007;**13**(6):CR270–4.

Marra 2002 *{published data only}*
Marra C, De Santo LS, Amarelli C, Della Corte A, Onorati F, Torella M, et al. Coronary artery bypass grafting in patients with severe left ventricular dysfunction: a prospective randomized study on the timing of perioperative intraaortic balloon pump support. *International Journal of Artificial Organs* 2002;**25**(2):141–6.

Moulopoulos 1986a *{published data only}*
Moulopoulos S, Stamatelopoulos S, Petrou P. Intraaortic balloon assistance in intractable cardiogenic shock. *European Heart Journal* 1986;**7**(5):396-403.

O'Neill 2012 *{published data only}*
O'Neill WW, Kleiman NS, Moses J, Henriques JPS, Dixon S, et al. A prospective, randomized clinical trial of hemodynamic support with impella 2.5 versus intra-aortic balloon pump in patients undergoing high-risk percutaneous coronary intervention: The PROTECT II study. *Circulation* 2012;**126**:1717–27.

O'Rourke 1981 *{published data only}*
O'Rourke MF, Norris RM, Campbell TJ, Chang VP, Sammel NL. Randomized controlled trial of intraaortic balloon counterpulsation in early myocardial infarction with acute heart failure. *American Journal of Cardiology* 1981;**47**:815–20.

Ohman 1994 *{published data only}*
* Ohman EM, George BS, White CJ, Kern MJ, Gurbel PA, Freedman RJ, et al. Use of aortic counterpulsation to improve sustained coronary artery patency during acute myocardial infarction. Results of a randomized trial. The Randomized IABP Study Group. *Circulation* 1994;**90**(2): 792–9.
Talley JD, Ohman EM, Mark DB, George BS, Leimberger JD, Berdan LG, et al. Economic implications of the prophylactic use of intraaortic balloon counterpulsation in the setting of acute myocardial infarction. The Randomized IABP Study Group. *American Journal of Cardiology* 1997; **79**(5):590–4.

Onorati 2005 *{published data only}*
Onorati F, Cristodoro L, Mastroroberto P, di Virgilio A, Esposito A, et al. Should we discontinue intraaortic balloon during cardioplegic arrest? Splanchnic function results of a prospective randomized trial. *Annals of Thoracic Surgery* 2005;**80**:2221–8.

Perera 2009 *{published data only}*
Perera D, Stables R, Booth J, Thomas M, Redwood S. The balloon pump-assisted coronary intervention study (BCIS-1): rationale and design. *American Heart Journal* 2009;**158** (6):910–6.

RECOVER II Trial *{published and unpublished data}*
NCT00972270. Trial Using Impella LP 2.5 System in Patients With Acute Myocardial Infarction Induced Hemodynamic Instability. http://www.controlled-trials.com/mrct/trial/629451/IABP (accessed 28 January 2010).

Sanborn 2000 *{published data only}*
Sanborn TA, Sleeper LA, Bates E, Jacobs AK, Boland J, French JK, et al. Impact of thrombolysis, intra-aortic balloon pump counterpulsation, and their combination in cardiogenic shock complicating acute myocardial infarction: a report from the SHOCK Trial Registry. SHould we emergently revascularize Occluded Coronaries for cardiogenic shocK?. *Journal of the American College of Cardiology* 2000;**36**:1123–9.

Stomel 1994 *{published data only}*
Stomel RJ, Rasak M, Bates ER. Treatment strategies for acute myocardial infarction complicated by cardiogenic shock in a community hospital. *Chest* 1994;**105**:997-1002.

Stone 1997 *{published data only}*
Stone GW, Marsalese D, Brodie BR, Griffin JJ, Donohue B, Costantini C, et al. A prospective, randomized evaluation of prophylactic intraaortic balloon counterpulsation in high risk patients with acute myocardial infarction treated with primary angioplasty. Second Primary Angioplasty in Myocardial Infarction (PAMI-II) Trial Investigators. *Journal of the American College of Cardiology* 1997;**29**(7):1459–67.

Stub 2011 *{published data only}*
Stub D, Chan W, Andrianopoulos N, Butler M, Clark DJ, et al. Comparison of percutaneous coronary intervention and coronary artery bypass grafting for revascularization of patients with diabetes mellitus. *Journal of the American College of Cardiology* 2011;**1**:E1895.

Taguchi 2000 *{published data only}*
Taguchi I, Ogawa K, Oida A, Abe S, Kaneko N, Sakio H. Comparison of hemodynamic effects of enhanced external counterpulsation and intra-aortic balloon pumping in patients with acute myocardial infarction. *The American Journal of Cardiology* 2000;**86**:1139–41.

Van 't Hof 1999 *{published data only}*
Van 't Hof AWJ, Liem AL, de Boer MJ, Hoorntje JCA, Suryapranata H, Zijlstra F. A randomized comparison of intra-aortic balloon pumping after primary coronary angioplasty in high risk patients with acute myocardial infarction. *European Heart Journal* 1999;**20**(9):659–65.

Vijayalakshmi 2007 *{published data only}*
Vijayalakshmi K, Kunadian B, Whittaker VJ, Wright RA, Hall JA, Sutton A, et al. Intra-aortic counterpulsation does not improve coronary flow early after PCI in a high-risk group of patients: observations from a randomized trial to explore its mode of action. *Journal of Invasive Cardiology* 2007;**19**(8):339–46.

Vis 2007a *{published data only}*
Vis MM, Sjauw KD, van der Schaaf RJ, Baan J Jr, Koch KT, DeVries JH, et al. In patients with ST-segment elevation

myocardial infarction with cardiogenic shock treated with percutaneous coronary intervention, admission glucose level is a strong independent predictor for 1-year mortality in patients without a prior diagnosis of diabetes. *American Heart Journal* 2007;**154**:1184-90.

Vis 2007b ***{published data only}***

Vis MM, Sjauw KD, van der Schaaf RJ, Koch KT, Baan J Jr, Tijssen JG, et al. Prognostic value of admission hemoglobin levels in ST-segment elevation myocardial infarction patients presenting with cardiogenic shock. *American Journal of Cardiology* 2007;**99**:1201-2.

Waksman 1993 ***{published data only}***

Waksman R, Weiss AT, Gotsman MS, Hasin Y. Intra-aortic balloon counterpulsation improves survival in cardiogenic shock complicating acute myocardial infarction. *European Heart Journal* 1993;**14**(1):71–4.

Zeymer 2011 ***{published data only}***

Zeymer U, Bauer T, Hamm C, Zahn R, Weidinger F, et al. Use and impact of intra-aortic balloon pump on mortality in patients with acute myocardial infarction complicated by cardiogenic shock: results of the Euro Heart Survey on PCI. *EuroIntervention* 2011;**7**:437–41.

Zeymer 2013 ***{published data only}***

Zeymer U, Hochadel M, Hauptmann KE, Wiegand K, Schuhmacher B, et al. Intra-aortic balloon pump in patients with acute myocardial infarction complicated by cardiogenic shock: results of the ALKK-PCI registry. *Clinical Research in Cardiology* 2013;**102**:223–7.

Additional references

Abdel-Wahab 2010

Abdel-Wahab M, Saad M, Kynast J, Geist V, Sherif MA, Richardt MA, et al. Comparison of hospital mortality with intra-aortic balloon counterpulsation insertion before versus after primary percutaneous coronary intervention for cardiogenic shock complicating acute myocardial infarction. *American Journal Cardiology* 2010;**105**(7):967–71.

AHA 2008

American Heart Association. *Heart Disease and Stroke Statistics - 2008 Update*. Dallas, Texas: American Heart Association, 2008.

Alcan 1983

Alcan KE, Stertzer SH, Wallsh E, DePasquale NP, Bruno MS. The role of intra-aortic balloon counterpulsation in patients undergoing percutaneous transluminal coronary angioplasty. *American Heart Journal* 1983;**105**(3):527–30.

Antman 2004

Antman EM, Anbe DT, Armstrong PW, Bates ER, Green LA, Hand M, et al. ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction--executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to revise the 1999 guidelines for the management of patients with acute myocardial infarction). *Journal of the American College of Cardiology* 2004;**44**(3):671–719.

Arafa 1999

Arafa OE, Pedersen TH, Svennevig JL, Fosse E, Geiran OR. Vascular complications of the intraaortic balloon pump in patients undergoing open heart operations: 15-year experience. *Annals of Thoracic Surgery* 1999;**67**:645–51.

Assis 2009

Assis RB, Azzolin K, Boaz M, Rabelo ER. Complications of intra-aortic balloon in a cohort of hospitalized patients: implications of nursing care. *Revista Latino-Americana de Enfermagem* 2009;**17**(5):658–63.

Bahekar 2012

Bahekar A, Singh M, Singh S, Bhuriya R, Khraisat A, et al. Cardiovascular outcomes using intra-aortic balloon pump in high-risk acute myocardial infarction with or without cardiogenic shock: a meta-analysis. *American Heart Journal* 2012;**164**:58–64.

BHF 2007

British Heart Foundation. *Coronary Heart Disease Statistics*. London: British Heart Foundation, 2007.

Cassese 2012

Cassese S, de Waha A, Ndrepepa G, Ranftl S, King L, et al. Intra-aortic balloon counterpulsation in patients with acute myocardial infarction without cardiogenic shock. A meta-analysis of randomized trials. *American Heart Journal* 2012; **164**:58–65.

Cheng 2009

Cheng JM, den Uil CA, Hoeks SE, van der Ent M, Jewbali LSD, van Domburg RT, et al. Percutaneous left ventricular assist devices vs. intra-aortic balloon pump counterpulsation for treatment of cardiogenic shock: a meta-analysis of controlled trials. *European Heart Journal* 2009;**30**(17):2102–8.

Cheng 2013

Cheng JM, van Leeuwen MAH, de Boer SPM, Tjong Joe Wai MCG, den Uil CA, et al. Impact of intra-aortic balloon pump support initiated before versus after primary percutaneous coronary intervention in patients with cardiogenic shock from acute myocardial infarction. *International Journal of Cardiology* 2013;**168**(4):3758–63.

Cohen 2003

Cohen M, Urban P, Christenson JT, Joseph DL, Freedman RJ Jr, Miller MF, et al. Contemporary utilization and outcomes of intraaortic balloon counterpulsation in acute myocardial infarction: the benchmark registry. *Journal of the American College of Cardiology* 2003;**24**(19):1763–70.

Cooper 2008

Cooper HA, Thompson E, Panza JA. The role of heparin anticoagulation during intra-aortic balloon counterpulsation in the coronary care unit. *Acute Cardiac Care* 2008;**10**(4): 214–20.

Debrunner 2008

Debrunner M, Schuiki E, Minder E, Straumann E, Naegeli B, Mury R, et al. Proinflammatory cytokines in acute myocardial infarction with and without cardiogenic shock. *Clinical Research in Cardiology* 2008;**97**(5):298–305.

Duchateau 2008
Duchateau L, Janssen P. *The Frailty Model*. Springer Science+Business Media, LLC, 2008.

Dyub 2008
Dyub AM, Whitlock RP, Abouzahr LL, Cinà CS. Preoperative intra-aortic balloon pump in patients undergoing coronary bypass surgery: a systematic review and metanalysis. *Journal of Cardiac Surgery* 2008;**23**(1): 79–86.

Erdogan 2006
Erdogan HB, Goksedef D, Erentug V, Polat A, Bozbuga N, Mansuroglu D, et al. In which patients should sheathless IABP be used? An analysis of vascular complications in 1211 cases. *Journal of Cardiac Surgery* 2006;**21**(4):342–6.

Forrester 1976a
Forrester JS, Diamond G, Chatterjee K, Swan HJ. Medical therapy of acute myocardial infarction by application of hemodynamic subsets (first of two parts). *New England Journal of Medicine* 1976;**295**(24):1356–62.

Forrester 1976b
Forrester JS, Diamond G, Chatterjee K, Swan HJ. Medical therapy of acute myocardial infarction by application of hemodynamic subsets (second of two parts). *New England Journal of Medicine* 1976;**295**(25):1404–13.

Forssell 1979
Forssell G, Nordlander R, Nyquist O, Schenk-Gustavsson K. Intraaortic balloon pumping in the treatment of cardiogenic shock complicating acute myocardial infarction. *Acta Medica Scandinavica* 1979;**206**(3):189–92.

Fuchs 2009
Fuchs S, Kornowski R, Teplitsky I, Brosh D, Lev E, Vaknin-Assa H, et al. Major bleeding complicating contemporary primary percutaneous coronary interventions ? incidence, predictors, and prognostic implications. *Cardiovascular Revascularization Medicine* 2009;**10**(2):88–93.

Gaziano 2010
Gaziano TA, Bitton A, Anand S, Abrahams-Gessel S, Murphy A. Growing epidemic of coronary heart disease in low- and middle-income countries. *Current Problems in Cardiology* 2010;**35**(2):72–115.

Geppert 2002
Geppert A, Steiner A, Zorn G, Delle-Karth G, Koreny M, Haumer M, et al. Multiple organ failure in patients with cardiogenic shock is associated with high plasma levels of interleukin-6. *Critical Care Medicine* 2002;**30**(9):1987–94.

Geppert 2006
Geppert A, Dorninger A, Delle-Karth G, Zorn G, Heinz G, Huber K. Plasma concentrations of interleukin-6, organ failure, vasopressor support, and successful coronary revascularization in predicting 30-day mortality of patients with cardiogenic shock complicating acute myocardial infarction. *Critical Care Medicine* 2006;**34**(8):2035–42.

Gjesdal 2009
Gjesdal O, Gude E, Arora S, Leivestad T, Andreassen AK, Gullestad L, et al. Intra-aortic balloon counterpulsation as a bridge to heart transplantation does not impair long-term survival. *European Journal of Heart Failure* 2009;**11** (7):709–14.

Goldberg 1999
Goldberg RJ, Samad NA, Yarzebski J, Gurwitz J, Bigelow C, Gore JM. Temporal trends in cardiogenic shock complicating acute myocardial infarction. *New England Journal of Medicine* 1999;**340**(15):1162–9.

Henriques 2006
Henriques JP, Remmelink M, Baan J Jr, van der Schaaf RJ, Vis MM, Koch KT, et al. Safety and feasibility of elective high-risk percutaneous coronary intervention procedures with left ventricular support of the Impella Recover LP 2.5. *American Journal of Cardiology* 2006;**97**:990-2.

Higgins 2002
Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 2002;**21**:1539–58.

Higgins 2011
Higgins JPT, Altman DG, Sterne JAC (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org. The Cochrane Collaboration.

Hochman 1999
Hochman JS, Sleeper LA, Webb JG, Sanborn TA, White HD, Talley JD, et al. Early revascularization in acute myocardial infarction complicated by cardiogenic shock. *New England Journal of Medicine* 1999;**341**(9):625–34.

Hochman 2001
Hochman JS, Sleeper LA, White HD, Dzavik V, Wong SC, Menon V, et al. One-year survival following early revascularization for cardiogenic shock. *JAMA* 2001;**285** (2):190–2.

Hochman 2003
Hochman JS. Cardiogenic shock complicating acute myocardial infarction: expanding the paradigm. *Circulation* 2003;**107**:2998–3002.

Hochman 2007
Hochman 2007, Apolito R. The calm after the storm. Long-term survival after cardiogenic shock. *Journal of the American College of Cardiology* 2007;**50**(18):1759–60.

Holmes 1997
Holmes DR Jr, Califf RM, Van de Werff F, Berger BP, Bates ER, Simoons ML, et al. Difference in countries' use of resources and clinical outcome for patients with cardiogenic shock after myocardial infarction: results from the GUSTO trial. *Lancet* 1997;**349**(9045):75–8.

Hudson 1999
Hudson MP, Granger CB, Stebbins AL. Cardiogenic shock survival and use of intraaortic balloon counterpulsation: results from the GUSTO I and III trials. *Circulation* 1999; **100 Suppl**:I370.

Ishihara 1991
Ishihara M, Sato H, Tateishi H, Uchida T, Dote K. Intraaortic balloon pumping as the postangiopplasty strategy in acute myocardial infarction. *American Heart Journal* 1991;**122**(2):385–9.

Kantrowitz 1968
Kantrowitz A, Tjonneland S, Freed PS, Phillips SJ, Butner AN, Sherman JL Jr. Initial clinical experience with intraaortic balloon pumping in cardiogenic shock. *JAMA* 1968;**203**:135–40.

Kocogullari 2008
Kocogullari CU, Emmiler M, Ayva E, Cekirdekct A. IABP-Related vascular complications: who is responsible--the patient, the surgeon or the sheath? Part I: sheath-related complications. *Advances in Therapy* 2008;**25**(3):225–30.

Kohsaka 2006
Kohsaka S, Menon V, Lowe AM, Lange M, Dzavik V, Sleeper LA, et al. Systemic inflammatory response syndrome after acute myocardial infarction complicated by cardiogenic shock. *Archives of Internal Medicine* 2005;**165** (14):1643–50.

Kontovannis 1999
Kontoyannis DA, Nanas JN, Kontoyannis SA, Stamatelopoulos SF, Moulopoulos SD. Mechanical ventilation in conjunction with the intra-aortic balloon pump improves the outcome of patients in profound cardiogenic shock. *Intensive Care Medicine* 1999;**25**(8): 835–8.

Kovack 1997b
Kovack PJ, Rasak MA, Bates ER, Ohman EM, Stomel RJ. Thrombolysis plus aortic counterpulsation: improved survival in patients who present to community hospitals with cardiogenic shock. *Journal of the American College of Cardiology* 1997;**29**(7):1454–8.

Kumbasar 1999
Kumbasar SD, Semiz E, Sancaktar O, Yalçinkaya S, De ğ er N. Mechanical complications of intra-aortic balloon counterpulsation. *International Journal of Cardiology* 1999; **70**(1):69–73.

Lewis 2007
Lewis PA, Mullany DV, Townsend S, Johnson J, Wood L, Courtney M, et al. Trends in intra-aortic balloon counterpulsation: comparison of 669 record Australian dataset with the multinational Benchmark Counterpulsation Outcomes Registry. *Anaesthesia and Intensive Care* 2007;**35** (1):13–9.

Lim 2003
Lim N, Dubois MJ, de Backer D, Vincent JL. Do all nonsurvivors of cardiogenic shock die with a low cardiac index?. *Chest* 2003;**124**:1885–91.

Lozano 2012
Lozano R, Naghavi M, Foreman K. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2010;**380**(9859):2095–128.

McEnany 1978
McEnany MT, Kay HR, Buckley MJ, Dagget WM, Erdmann AJ, Mundth ED, et al. Clinical experience with intraaortic balloon pump support in 728 patients. *Circulation* 1978;**58**:I124–32.

Menon 2000
Menon V, White H, LeJemtel T, Webb JG, Sleeper LA, Hochman JS. The clinical profile of patients with suspected cardiogenic shock due to predominant left ventricular failure: a report from the SHOCK Trial Registry. SHould we emergently revascularize Occluded Coronaries in cardiogenic shocK?. *Journal of the American College of Cardiology* 2000;**36**(3 Suppl A):1071–6.

Moher 2009
Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine* 2009;**6**(7):1–6.

Moran 2014
Moran AE, Forouzanfar MH, Roth GA, Mensah GA, Ezzati M, Murray CJ, et al. Temporal trends in ischemic heart disease mortality in 21 world regions, 1980 to 2010: the Global Burden of Disease 2010 study. *Circulation* 2014; **129**(14):1483–92.

Moulopoulos 1986b
Moulopoulos S, Stametolopoulos S, Petrou P. Intraaortic balloon assistance in intractable cardiogenic shock. *European Heart Journal* 1986;**7**(5):396–403.

Mozaffarian 2014
Mozaffarian D, Fahimi S, Singh GM, Micha R, Khatibzadeh S, Engell RE, et al. Global Burden of Diseases Nutrition and Chronic Diseases Expert Group. Global sodium consumption and death from cardiovascular causes. *New England Journal of Medicine* 2014;**371**(7):624–34.

Nieuwlaat 2013
Nieuwlaat R, Schwalm JD, Khatib R, Yusuf S. Why are we failing to implement effective therapies in cardiovascular disease?. *European Heart Journal* 2013;**34**(17):1262–9.

O'Gara 2013
O'Gara PT, Kushner FG, Ascheim DD, Casey DE Jr, Chung MK, et al. 2013 ACCF/AHA Guideline for the Management of ST-Elevation Myocardial Infarction: A Report of the American College of Cardiology Foundation/ American Heart Association Task Force on Practice Guidelines. *Circulation* 2013;**127**(4):529–55. [DOI: 10.1161/CIR.0b013e3182742c84; :]

Ohman 2001
Ohman EM, Hochmann JS. Aortic counterpulsation in acute myocardial infarction: physiologically important but does the patient benefit?. *American Heart Journal* 2001;**141**: 889–92.

Ohman 2005a
Ohman EM, Nanas J, Stomel RJ, Leesar MA, Nielsen DW, O'Dea D, et al. Thrombolysis and counterpulsation to improve survival in myocardial infarction complicated by hypotension and suspected cardiogenic shock or heart

failure: results of the TACTICS trial. *Journal of Thrombosis and Thrombolysis* 2005;**19**(1):33–9.

Pfeiffer 2005
Pfeiffer S, Frisch P, Weyand M, Ensminger SM, Tandler R, Sirch J, et al. The use of preoperative intra-aortic balloon pump in open heart surgery. *Journal of Cardiovascular Surgery (Torino)* 2005;**46**(1):55–60.

Piedbois 2004
Piedbois P, Mondor H, Buyse M. Meta-analyses based on abstracted data: A step in the right direction, but only a first step. *Journal of Clinical Oncology* 2004;**22**(19):3839–41.

Ramnarine 2005
Ramnarine IR, Grayson AD, Dihmis WC, Mediratta NK, Fabri BM, Chalmers JA. Timing of intra-aortic balloon pump support and 1-year survival. *European Journal of Cardiothorac Surgery* 2005;**27**(5):887–92.

Riaz 2008
Riaz W, Shahbaz A, Sami W, Khan JS. Early vascular complications of intraaortic balloon counterpulsation in patients undergoing open heart surgery. *Journal of Ayub Medical College, Abbottabad* 2008;**20**(2):80–4.

Riley 2007
Riley RD, Simmonds MC, Look MP. Evidence combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. *Journal of Clinical Epidemiology* 2007;**60**(5):431–9.

Riley 2010
Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;**340**(c221):521–5.

Romeo 2013
Romeo F, Acconcia MC, Sergi D, Romeo A, Muscoli S, et al. The outcome of intra-aortic balloon pump support in acute myocardial infarction complicated by cardiogenic shock according to the type of revascularization: a comprehensive meta-analysis. *American Heart Journal* 2013;**165**:679–92.

Santa-Cruz 2006
Santa-Cruz RA, Cohen MG, Ohman EM. Aortic counterpulsation: A review of the hemodynamic effects and indications for use. *Catheterization & Cardiovascular Interventions* 2006;**67**(1):68–77.

Seely 2000
Seely AJ, Christou NV. Multiple organ dysfunction syndrome: Exploring the paradigm of complex nonlinear systems. *Critical Care Medicine* 2000;**28**(7):2193–200.

Sjauw 2009
Sjauw KD, Engström AE, Vis MM, van der Schaaf RJ, Baan J, Koch KT, et al. A systematic review and meta-analysis of intra aortic balloon pump therapy in ST-elevation myocardial infarction: should we change the guidelines?. *European Heart Journal* 2009;**30**(4):459–68.

Sleeper 2005
Sleeper LA, Ramanathan K, Picard MH, Lejemtel TH, White HD, et al. Functional status and quality of life after emergency revascularization for cardiogenic shock complicating acute myocardial infarction. *Journal of the American College of Cardiology* 2005;**46**(2):266–73.

Spencer 2001
Spencer FA, Goldberg RJ, Frederick PD, Malmgren J, Becker RC, Gore JM. Age and the utilization of cardiac catheterization following uncomplicated first acute myocardial infarction treated with thrombolytic therapy (The Second National Registry of Myocardial Infarction [NRMI-2]). *The American Journal of Cardiology* 2001;**88**(2):107–11.

Steg 2012
Steg G, James SK, Atar D, Badano LP, Blomstrom Lundqvist C, et al. ESC GUIDELINES: ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force on the management of ST-segment elevation acute myocardial infarction of the European Society of Cardiology (ESC). *European Heart Journal* 2012;**33**(20):2569–619. [DOI: doi:10.1093/eurheartj/ehs215]

Stone 2003
Stone GW, Ohman EM, Miller MF, Joseph DL, Christenson JT, Cohen M, et al. Contemporary utilization and outcomes of intraaortic balloon counterpulsation in acute myocardial infarction: the benchmark registry. *Journal of the American College of Cardiology* 2003;**41**(11):1940–5.

Takano 1984
Takano T, Endo T, Saito H, Ida T, Tanaka K, Osaka S, et al. Clinical usefulness of intraaortic balloon pumping in acute myocardial infarction complicated with cardiogenic shock, ventricular septal perforation and mitral regurgitation. *Japanese Circulation Journal* 1984;**48**(7):678–89.

The Global Burden Collaboration 2014
The Global Burden of Metabolic Risk Factors for Chronic Diseases Collaboration. Cardiovascular disease, chronic kidney disease, and diabetes mortality burden of cardiometabolic risk factors from 1980 to 2010: a comparative risk assessment. *The Lancet Diabetes & Endocrinology* 2014;**2**(8):634–47.

Theologou 2011
Theologou T, Bashir M, Rengarajan A, Khan O, Spyt T, Richens D, Field M. Preoperative intra aortic balloon pumps in patients undergoing coronary artery bypass grafting. *Cochrane Database of Systematic Reviews* 2011, Issue 1. [DOI: 10.1002/14651858.CD004472.pub3]

Thiele 2001
Thiele H, Lauer B, Hambrecht R, Boudriot E, Cohen HA, Schuler G. Reversal of cardiogenic shock by percutaneous left atrial-to-femoral arterial bypass assistance. *Circulation* 2001;**104**:2917-22.

van de Werf 2008
Van de Werf F, Bax J, Betriu A, Blomstrom-Lundqvist C, Crea F, Falk V, et al. Management of acute myocardial infarction in patients presenting with persistent ST-segment elevation: the Task Force on the Management of ST-

Segment Elevation Acute Myocardial Infarction of the European Society of Cardiology. *European Heart Journal* 2008;**29**(23):2909–45.

Weiss 1984
Weiss AT, Engel S, Gotsman CJ, Shefer A, Hasin Y, Bitran D, et al. Regional and global left ventricular function during intra-aortic balloon counterpulsation in patients with acute myocardial infarction shock. *American Heart Journal* 1984;**108**(2):249–54.

Werdan 2010
Werdan K, Ruß M, Buerke M. Procedures in ICU: IABP. In: Tubaro M, Danchin N, Filippatos G, Goldstein P, Vranckx P, Zahger D editor(s). *ESC textbook of intensive and acute cardiac care*. England: Oxford University Press, 2010.

Werdan 2012
Werdan K, Ruß M, Buerke M, Delle-Karth G, Geppert A, Schondube FA, et al. Cardiogenic shock due to myocardial infarction: diagnosis, monitoring and treatment: a German-Austrian S3 Guideline. *Deutsches Arzteblatt International* 2012;**109**(19):343–51.

Werdan 2014
Werdan K, Gielen S, Ebelt H, Hochman JS. Mechanical circulation support in cardiogenic shock. *European Heart Journal* 2014;**35**(3):156–67.

Whitehead 2002
Whitehead N. *Meta-analysis of controlled clinical trials*. Chichester: John Wiley & Sons, 2002.

Yusuf 2004
Yusuf S, Hawken S, unpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 2004; **364**:937–52.

Yusuf 2006
Yusuf S, Mehta SR, Chrolavicius S, Afzal R, Pogue J, Granger CB, et al. Comparison of fondaparinux and enoxaparin in acute coronary syndromes. *New England Journal of Medicine* 2006;**354**(14):1464–76.

Zhang 2013
Zhang XR, Su L, Chen SJ, Zhang XG, Ying YH. A meta-analysis on efficiency of intra-aortic balloon pump therapy in patients with acute myocardial infarction. *Chinese Journal of Cardiovascular Disease (Zhonghua Xin Xue Guan Bing Za Zhi)* 2013;**41**(5):432–7.

* *Indicates the major publication for the study*

Unverzagt S, Prondzinsky R, Peinemann F (2013) Single-center trials tend to provide larger treatment effects than multicenter trials: a systematic review. J Clin Epidemiol 66: 1271-80.

ELSEVIER

Journal of Clinical Epidemiology 66 (2013) 1271−1280

Journal of Clinical Epidemiology

REVIEW ARTICLE

Single-center trials tend to provide larger treatment effects than multicenter trials: a systematic review

Susanne Unverzagt[a,*], Roland Prondzinsky[b], Frank Peinemann[c]

[a]Institute of Medical Epidemiology, Biostatistics and Informatics, Martin Luther University of Halle-Wittenberg, 06097 Halle (Saale), Germany
[b]Department of Medicine III, Martin Luther University of Halle-Wittenberg, 06097 Halle (Saale), Germany
[c]Children's Hospital, University of Cologne, Cologne, Germany

Accepted 14 May 2013; Published online 20 August 2013

Abstract

Objectives: To assess whether the reported trial characteristics are associated with treatment effects on all-cause mortality within critical care medicine.

Study Design and Setting: We identified all eligible randomized controlled trials (RCTs) from Cochrane Reviews on patients with sepsis, septic shock, and cardiogenic shock. Risk of bias was judged on 12 trial characteristics, including the differentiation between single-center and multicenter trials. Hierarchical random-effects models quantified the impact of the risk of bias items on the reported effect estimates of mortality.

Results: Twelve meta-analyses that involved 82 RCTs were selected and judged. Single-center trials estimated a significant larger treatment effect compared with multicenter trials (ratio of odds ratios, 0.64; 95% confidence interval: 0.47, 0.87). Treatment effect tended to be overestimated with selective reporting of preplanned end points. Biases in different trial characteristics are unlikely to operate independently and may have modified these associations.

Conclusion: The results of this study highlight a substantial difference in treatment effect estimates between single-center and multicenter trials. Therefore, we recommend that results from single-center trials should be cautiously used for decision making.

Keywords: Risk of bias; Randomized controlled trials; Research design; Confounding factors; Risk of bias; Critical care medicine

1. Introduction

Randomized controlled trials (RCTs), when appropriately designed, conducted, and reported, provide the most valid assessment of treatment effects [1,2]. Furthermore, systematic reviews of RCTs are considered as the most comprehensive way of judging whether a treatment ''does more good than harm'' [3].

Hence, defects in the trial characteristics included in a systematic review can have a substantial impact on estimates, influence the validity of conclusions of a review [4,5], and may compromise recommendations in clinical guidelines and therefore the quality of health care received by patients. For this reason, careful appraisal of these trial characteristics in primary studies is an essential feature of systematic reviews and helps to identify strengths and weaknesses in the existing evidence [6].

Empirical evidence on specific trial characteristics associated with bias in estimated treatment effects of meta-analyses has come from collections of meta-analyses in so called meta-epidemiological trials [7]. Differences in the risk of bias may indicate that effect estimates of some trials are more biased than others.

It is assumed that the relevance of several trial characteristics strongly depends on the investigated clinical topic [5]. Our study concentrates on reviews within emergency and critical care medicine and investigates the outcome of all-cause mortality to minimize heterogeneity between meta-analyses. Randomized trials concerning patients with acute and partially rare disease are regularly small and heterogeneous because of organizational, ethical, and judicial problems. Furthermore, intensive care varies between and even within countries and is determined by resources, case mix, and the culture of end-of-life care. Bellomo et al. [8] have highlighted a number of examples of single-center trials in critical care medicine with substantial shortcomings and

Conflict of interest: The authors declare that they have no competing interests.

* Corresponding author. Tel.: +49-345-557-4486; fax: +49-345-557-3580.

E-mail address: susanne.unverzagt@medizin.uni-halle.de (S. Unverzagt).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2013.05.016

What is new?

Key findings

- Single-center trials in critical care medicine tend to provide larger treatment effect estimates than multicenter trials.
- Randomized controlled trial (RCT) characteristics did not show a strong association with outcome of all-cause mortality in other investigated characteristics including sequence generation, allocation concealment, double blinding, incomplete outcome data, selective outcome reporting, early stopping, preinterventions, competing interests, baseline imbalances, crossover, and short follow-up periods.
- Future studies examining the association between trial characteristics and treatment effect should control the number of study centers.

What this adds to what was known?

- In critical care medicine, evidence from single-center trials seems to overestimate treatment effects, but larger multicenter RCTs are much more difficult to conduct in the setting of the intensive care unit in regard to comparability of baseline conditions and also financing and sponsoring.

What is the implication and what should change now?

- There is a need for critical judgment of the results of single-center trials. If single-center trials are used for decision making, their results should be used with caution.
- RCTs investigating expensive therapeutic concepts should be financed and supervised by national and international scientific societies and not solely by pharmaceutical companies.

gave different causes for an overestimation of treatment effects.

The aim of the present study was to assess the internal validity of the existing evidence. Furthermore, we estimated the consequences of bias in 12 potential trial characteristics from the literature to the estimated treatment effect on all-cause mortality.

2. Methods

2.1. Selection of meta-analyses

We searched systematic reviews on critically ill patients with indications of cardiogenic shock, sepsis, severe sepsis, and septic shock in The Cochrane Library. Patients with these indications represent a high proportion of those in intensive care units. Because of the extent of underlying cardiovascular diseases and demographical changes, an increased incidence and mortality of these indications are expected in the future [9]. We included all interventions to increase blood pressure, maintain adequate organ and tissue perfusion, and prevent patients from multiorgan failure according to the fundamental measures of early goal-directed therapy. This included volume resuscitation or substitution, vasoactive or inotropic therapy, anti-infective therapy, correction of hematocrit by transfusion of red cell packages, including immunoglobulin administration, and blood glucose control and regulation. All-cause mortality was chosen as the primary outcome. In fact, it is regularly reported in the investigated clinical area and represents the ultimate clinical benefit for the patient, provided that quality of life is not compromised.

We searched in the Cochrane Database of Systematic Reviews (2011, issue 11) on patients with cardiogenic shock, severe sepsis, and septic shock in title, keyword, or abstract [10].

Two authors independently screened these reviews for eligible meta-analyses using the following criteria: at least three full-article RCTs that had been pooled and all-cause mortality was meta-analyzed.

2.2. Quality assessment

Only few studies describe trial characteristics in critical care or emergency medicine [8,11−13]. We identified 12 characteristics previously demonstrated or hypothesized to be associated with treatment effect estimates by reviewing the Cochrane handbook [14] and published meta-epidemiological studies [7,11,13,15−26].

Two reviewers (S.U. and R.P.) independently evaluated a pilot testing of the 12 trial characteristics with a sample of 15 trials that are representative of the investigated clinical area. In a preliminary evaluation, we developed detailed instructions on judgment of the potential risk of bias on the basis of The Cochrane Collaboration risk of bias tool [14] and suggestions made by Hartling et al. [27] (double blinding, incomplete outcome data addressed, selective outcome reporting, and competing interests), Tierney and Stewart [26] and Fergusson et al. [28] (incomplete outcome data addressed), Balk et al. [11] (early stopping), Gluud [4] (competing interests), van Nieuwenhoven et al. [13] (baseline imbalances), and Rosén [29] (sufficient length of follow-up). The resulting instructions on judgments are summarized in Table 1.

On the basis of the selected trial characteristics and instructions given in Table 1, two reviewers (S.U. and F.P.) applied this risk of bias tool on the remaining 67 trials. The reviewers were not blinded to the study results. We discussed all unclear decisions until consensus was established.

Table 1. Criteria for judging risk of bias on 11 trial characteristics

Trial characteristics	Risk of bias	Definition
Sequence generation	High	Sequence generation based on a date or number; allocation by judgment of the clinician or preference of the participant, on the results of a laboratory test, a series of tests, or availability of the intervention
	Low	Use of random number table or generator; minimization
Allocation concealment	High	Open random allocation schedule; assignment envelopes without appropriate safeguards; alternation or rotation; date of birth and case record number; any other explicitly unconcealed procedure
	Low	Central allocation including telephonic, web-based, and pharmacy-controlled randomization; sequentially numbered drug containers of identical appearance and opaque and sealed envelopes
Double blinding	High	No blinding, incomplete blinding, or absence of placebo; blinding could have been broken
	Low	Blinding of participants and health-care providers, unlikely that the blinding could have been broken; nonblinding unlikely to introduce bias; method with placebo(s) or dummy technique
Incomplete outcome data addressed	High	Outcome-related reasons for missing outcome data; clinically relevant proportion of missing outcomes; only disease-related mortality or ''as-treated'' analysis
	Low	Reasons for missing outcome data unlikely to be related to true outcome and ineligibility was detected blinded to assignment and outcome; number and reasons of missing outcome data balanced across intervention groups; small proportion of missing outcomes ($<10\%$); imputation with appropriate methods
Selective outcome reporting	High	Missing prespecified outcomes; incomplete reporting of measurements, analysis methods, subsets, or time points of the data; missing information to patient-relevant outcomes in critically ill patients as mortality or adverse events.
	Low	All prespecified outcomes reported in a prespecified way as described in the protocol, methods, or objectives
Early stopping	High	Early stopping because of some data-dependent process without adjustments of alpha error and unplanned decisions of the data safety committee (frequency of harms); attained sample size is much less than the intended size or no reporting of a prespecified sample size; stopping to reasons that may be related to the observed intervention effect (lower than expected recruitment rate, no supply of drug or further financial support, and newly published results of other trials)
	Low	No early stopping (comparison with preplanned sample size calculation or explicit reporting required); preplanned early analyses with adjustments of alpha error
Preintervention	High	An intervention before randomization could influence the effect of the randomized intervention
	Low	Neither contra-active nor similar supporting prerandomization intervention were allowed; patients with these preintervention were excluded
Competing interests	High	One of the authors or a sponsor might financially benefit from the publication of the article and insufficient safeguards to guard against bias
	Low	No industry-initiated trial; inappropriate influence of the study sponsor can be ruled out by independent monitoring, audit, or analyses described; judgment should be blinded to study results
Baseline imbalances	High	Groups noncomparable in more than one factor strongly related to mortality (differences $<10\%$) as age, gender, acute physiology, and chronic health evaluation scores; hemodynamics, infectious profile, underlying disease, main comorbidities, information describing multiple organ dysfunction score, and inflammation
	Low	Groups are comparable in these factors
Crossover	High	Reason for cross-over data likely to be related to true outcome (e.g., decision of health-care providers) with imbalance in either numbers or reasons between intervention groups (differences of proportions $>10\%$); as-treated analysis done with substantial departure of the intervention received from that assigned at randomization
	Low	No crossover or low proportion $<10\%$; reasons for crossover unlikely to be related to true outcome (decision of the patient not related to health-care provider because of protocol violation); crossover balanced in numbers across intervention groups, with similar reasons across groups
Sufficient follow-up	High	Insufficient short follow-up period as until discharge from intensive care unit
	Low	All-cause mortality is reported with a minimal follow-up period until discharge from hospital, total recovery or over at least 28 days

2.3. Data extraction

All available information was extracted from the original trial reports and added by unpublished information reported in the systematic Cochrane reviews. To estimate treatment effect, information from a two-by-two table summarizing the number of deaths in each treatment group and the total number of patients randomly assigned to each group was extracted and compared with that from the review. In case of missing in-hospital mortality, 30- or 28-day mortalities or mortality rates over other follow-up periods were used.

In addition to the characteristics given in Table 1 and the treatment effect of each randomized trial, we extracted general characteristics of the trial: the publishing journal, date of publication, and sample size. We recorded whether the trial was a single-center or a multicenter (defined as more than one different centers) trial.

2.4. Statistical analyses

Before we examined associations between treatment effect and quality components, we estimated treatment effects as odds ratios (OR). Outcome was recoded so that an OR $<$ 1 indicated a beneficial effect of the experimental intervention. Mantel−Haenszel random-effects meta-analyses were used to combine intervention effects across trials within each meta-analysis. We calculated a fixed- and a random-effects multilevel model to assess the bias associated with trial characteristics in a sample of trials from different meta-analyses. Two statistical models were used to assess the bias associated with trial characteristics in a sample of trials and meta-analyses. Both models use the numbers of death in the two treatment arms of the trials to recalculate individual patient data. These data were used to predict the probability of a death for an individual patient in one of the included trials on the basis of treatment group, meta-analysis, judgment on trial characteristics, and interactions between judgment and treatment and between treatment and meta-analysis.

The first is a standard logistic regression model as described by Sterne et al. [30] and used by Siersma et al. [31]. This fixed-effects multilevel model was used to calculate starting values for the second random-effects model.

This multivariate random-effects approach predicts and explains some of the bias mechanism between co-occurring bias associated with trial characteristics and incorporated heterogeneity between trials into the model [31,32].

The results of the random-effects model with adjustments for 12 trial characteristics are interpreted as the primary analysis. The results of a fixed-effects model and models with adjustments to a single characteristic are interpreted as sensitivity analyses. This bias attributed to a special characteristic on the OR can be expressed as a ratio of odds ratios (ROR) comparing ORs of studies at high or unclear risk of bias with those of studies at low risk of bias.

An ROR <1 indicates an overestimation of treatment effect with larger ORs in low-risk trials compared with those in unclear or high-risk trials. An ROR >1 indicates an underestimation of treatment effect.

However, models adjust for the correlation between patients from the same trial. Correlation is assumed to be identical for every two patients from the same trial, and patients from different trials are considered to be independent. Interactions are included to investigate meta-analysis as an effect modifier. Analyses have been done in RevMan [33], PASW 18 [34], and SAS 9.2 (PROC LOGISTIC statement and PROC NLMIXED statement) [35]. Heterogeneity of the results of single meta-analyses was quantified using the I^2 value [36].

3. Results

3.1. Characteristics of included meta-analyses and trials

Meta-analyses were found by searching the Cochrane Database of Systematic Reviews (2011, issue 1) for ‘‘shock’’ and ‘‘sepsis.’’ We found 44 reviews and excluded 38. The selection process was recorded in a flow chart shown in Fig. 1.

Six eligible systematic reviews (Alejandria et al. [37], Annane et al. [38], Boeuf et al. [39], Havel et al. [40], Marti-Carvajal et al. [41], and Unverzagt et al. [42]) were selected, of which 82 RCTs could be pooled in 12 meta-analyses concerning all-cause mortality. A total of 24,657 patients were involved in these investigations (Table 2). Reviews were published between 2002 and 2011 in the Cochrane Database of Systematic Reviews. Included trials were published between 1956 and 2010 in 33 journals.

Table 2 presents detailed characteristics for each meta-analysis. Reviews include patients with sepsis, severe sepsis, or septic, cardiogenic, hemorrhagic, or spinal shock. Reviews examined the effects of intravenous immunoglobulin, corticosteroids, naloxone, vasopressors, human recombinant activated protein C, and intra-aortic balloon pump counterpulsation on all-cause mortality.

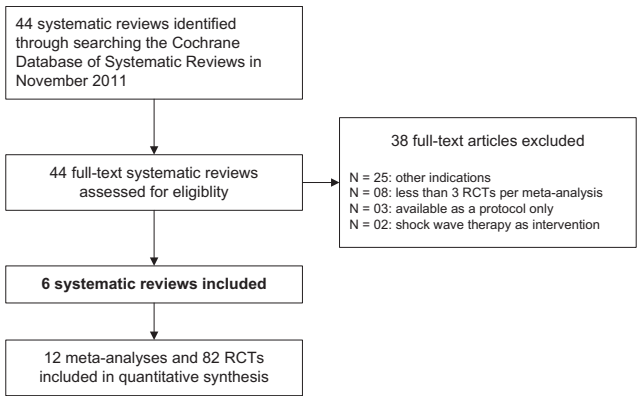


Fig. 1. Study flow chart. N, number; RCT, randomized controlled trial.

Table 2. Characteristics of included systematic reviews

Systematic review	Condition studied	Intervention in experimental group	Intervention in control group	Meta-analysis, *n*	RCTs per meta-analysis, *n* single-/multicenter	Patients: total; median (min–max)	All-cause mortality, OR (95% CI), I^2
Alejandria et al. [37]	Adults, children, and neonates (with sepsis or septic shock caused by bacteria	Monoclonal immunoglobulins (anticytokines or antiendotoxins), standard, or IgM-enriched polyclonal intravenous immunoglobulins	Placebo or no intervention and no immunoglobulins	4	9/0; 6/2; 4/11; 2/2	7,894; 696 (42–2,634) 4,673; 292.5 (33–2,199) 1,493; 52 (21–624) 192; 50.5 (31–60)	0.87 (0.79, 0.95), 0% 0.86 (0.66, 1.12), 63% 0.52 (0.35, 0.76), 44% 0.62 (0.30, 1.28), 0%
Annane et al. [38]	Children and adults with severe sepsis or septic shock and patients with other severe infections	Any intravenous treatment with corticosteroid preparations	Placebo or standard therapy	1	9/10	2,356; 59 (29–499)	0.73 (0.54, 0.99), 49%
Boeuf et al. [39]	Patients with septic, cardiogenic, hemorrhagic, or spinal shock	Naloxone	Placebo	1	0/3	64; 22 (14–28)	0.75 (0.21, 2.67), 24%
Havel et al. [40]	Acutely and critically ill adult and pediatric patient (no preterm infants)	Vasopressors (norepinephrine, epinephrine, and vasopressin)	Different vasopressors, intravenous fluids, and placebo	4	1/5; 2/1; 1/2; 1/3	2,043; 41 (10–1,679) 819; 23 (18–778) 382; 30 (22–330) 157; 39 (10–69)	0.82 (0.69, 0.98), 0% 1.18 (0.88, 1.57), 0% 1.14 (0.76, 1.70), 0% 1.02 (0.43, 2.41), 23%
Marti-Carvajal et al. [41]	Adults and children with severe sepsis or septic shock	Recombinant human activated protein C	Placebo	1	0/3	4,434; 1,690 (131–2,613)	0.89 (0.63, 1.24), 74%
Unverzagt et al. [42]	Adult patients with a clinical diagnosis of myocardial infarction complicated by cardiogenic shock	IABP counterpulsation	Non-IABP standard treatment and other assisting devices	1	3/2	150; 26 (21–41)	1.02 (0.53, 1.98), 0%

Abbreviations: RCT, randomized controlled trial; min, minimum; max, maximum; I^2, statistic for quantifying inconsistency; OR, odds ratio; CI, confidence interval; IABP, intra-aortic balloon pump.

All-cause mortality OR: combined random effects.

The included trials show a high variability in sample size ranging from 10 to 2,634 randomized patients. The combined OR from individual random-effects meta-analyses ranged from 0.52 to 1.18. Only two meta-analyses showed substantial heterogeneity ($I^2 > 60\%$). Eight of 12 meta-analyses favored the experimental intervention.

Among the 82 trials included in this study, 41 (50%) trials were performed as single-center trials and 41 (50%) as multicenter trials. Single-center and multicenter trials differed substantially in sample size: median size was 40 (range, 10−252) for single-center trials and 223 (range, 21−2,634) for multicenter trials.

3.2. Trial characteristics

A summary of the review author's judgments about each risk of bias item presented as percentages across all included studies is shown in the risk of bias graph in Fig. 2. We used the Cochrane risk of bias tool [43]. We found that single-center trials are associated with higher risk of bias or unclear reporting in nearly all characteristics (data not shown). Only inadequate outcome reporting of all-cause mortality is similarly distributed in single-center and multicenter trials.

3.3. Association between trial characteristics and estimated treatment effects

In general, multicenter trials showed higher frequencies of trials with low risk of bias. The percentage of trials with low risk of bias varied between 34% in trial characteristic early stopping and 84% in the reporting of incomplete outcome data. Characteristics from The Cochrane Collaboration's risk of bias tool were more frequently judged as low risk of bias compared with the indication-specific characteristics.

Table 3 summarizes RORs on the investigated association between risk of bias and treatment effect estimate of all investigated trial characteristics. Point estimates for RORs of high- or unclear risk vs. low-risk trials range from 0.64 to 1.13. However, only one of the investigated 12 characteristics was found to be significantly associated with treatment effect. Trends toward association were observed for further seven characteristics. Overall, estimated treatment effect was larger for single-center trials than for multicenter trials with an overestimation of 36% [ROR, 0.64; 95% confidence interval (CI): 0.47, 0.87]. This exaggeration is consistent over primary (Table 3) and sensitivity models (Appendix at www.jclinepi.com).

Selective outcome is another characteristic in which trials with high or unclear risk of bias might be related with a mean overestimation of 20% of treatment effect. This is reflected by a low ROR in this characteristic (ROR, 0.80; 95% CI: 0.57, 1.12). Only slightly exaggerated estimates of treatment effects with broad confidence intervals reflecting high heterogeneity might be related to contra-active or similar supporting interventions before randomization (ROR, 0.86; 95% CI: 0.73, 1.00), crossover (ROR, 0.89; 95% CI: 0.61, 1.31), and baseline imbalances (ROR, 0.90; 95% CI: 0.75, 1.07).

Shrinking of the observed treatment effects was found in trials with inadequate allocation concealment or incomplete outcome data. Treatment effect was relatively 13% smaller with broad CIs in trials with high or unclear risk of bias in

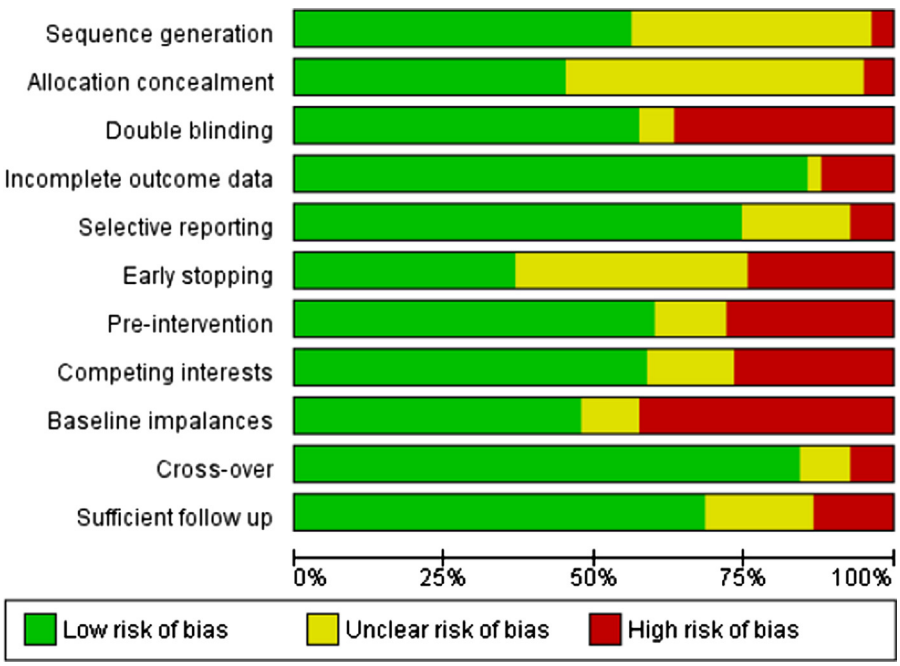


Fig. 2. Risk of bias graph: review author's judgments about each risk of bias item presented as percentages across all included studies. We used the Cochrane risk of bias tool to generate this figure.

Table 3. Association of risk of bias in 12 trial characteristics on the treatment effect: primary analysis

Trial characteristics	All 12 variables combined, random-effects model
Sequence generation	0.97 (0.76, 1.24)
Allocation concealment	1.13 (0.94, 1.19)
Double blinding	1.03 (0.80, 1.32)
Incomplete outcome data addressed	1.13 (0.89, 1.42)
Selective outcome reporting	0.80 (0.57, 1.12)
Early stopping	1.10 (0.94, 1.29)
Preintervention	0.86 (0.73, 1.00)
Competing interests	1.05 (0.88, 1.24)
Baseline imbalances	0.90 (0.75, 1.07)
Crossover	0.89 (0.61, 1.31)
Sufficient follow-up	1.11 (0.90, 1.38)
Multi vs. single-center trials	0.64 (0.47, 0.87)
Variance of between-study heterogeneity	0.43 (0.25, 0.61)

Abbreviation: ROR, ratio of odds ratios.

RORs estimated by logistic regression models on recalculated individual patient data to investigate the influence of risk of bias in 12 trial characteristics on the treatment effect. We calculated a ROR and used the odds ratio of high or unclear risk of bias as the numerator and used that odds ratio of low risk of bias as the denominator (95% confidence interval).

allocation concealment (ROR, 1.13; 95% CI: 0.94, 1.19) or that missing information on all-cause mortality was inadequately addressed compared with that of low-risk trials (ROR, 1.13; 95% CI: 0.89, 1.42).

No association was found between risk of bias in characteristics concerning sequence generation, double blinding, early stopping, competing interests, or insufficient follow-up and treatment effect estimates on all-cause mortality.

Comparisons between models controlling for all trial characteristics and models controlling for one single characteristic in which each bias coefficient is assessed separately show comparable results in most of the investigated characteristics. Confounding between characteristics may have modified RORs calculated by different models controlling for one single and 12 characteristics. Differences with completely different interpretations were found in the influence of double blinding and crossover. High-bivariate associations may explain occurrence of indications of confounding in the present sample.

Clear evidence for a confounding effect because of risk of bias in other trial characteristics was found in the case of insufficient blinding. The estimated ROR changed from 0.84 (95% CI: 0.69, 1.02) in the model with one single characteristic to 1.03 (95% CI: 0.80, 1.32) with all characteristics. Furthermore, RORs to describe the modification of estimated treatment effect by trials with crossover increased from 0.68 (95% CI: 0.49, 0.97) to 0.89 (95% CI: 0.61, 1.31), respectively.

4. Discussion

This study is based on an analysis of 82 RCTs in emergency and critical care medicine, adds six trial characteristics with detailed instructions on how to judge risk of bias to The Cochrane Collaboration's tool, and investigates the influence of single-center and multicenter trials. These influences are quantified by RORs. To prohibit overemphasized results on investigations, all investigated characteristics were prespecified before judgments, and summarizing statistics of all investigated associations are shown and discussed.

RORs were <1 for five trial characteristics, >1 for two trial characteristics, and indicated no difference for the remaining five trial characteristics. Overoptimistic estimates of treatment effects, associated with low ROR′, were found in single-center trials. Single-center trials showed significantly larger treatment effects than multicenter trials. The amount of the difference was substantial with a relative overestimation of 36% in the treatment effect for single-center trials compared with that for multicenter trials. Several reasons may explain this overestimation. The small-study effect in studies with a sample size that is too small to detect a clinically plausible effect (Sterne et al. [30]), publication bias (von Elm et al. [44]), a more homogeneous population, or a higher risk of bias compared with multicenter trials. Our data confirm a higher risk of bias of single-center trials in nearly all investigated characteristics. No difference was found in the incomplete outcome data, which may be explained by logistical issues.

Selective outcome reporting, contra-active or similar supporting interventions prerandomization, crossover, and baseline imbalances were associated with exaggerated estimates of treatment effect of relative 20%, 14%, 11%, and 10%. The influence of these four characteristics was stated by models with adjustments for one single or all characteristics.

The leading paradigm in the field of quality assessment of randomized trials is that low-quality trials tend to overestimate effect estimates [19,21,23,25]. However, different meta-epidemiological studies have shown that trial characteristics can indeed influence effect estimates in both directions [11,15]. In the present study, characteristics were identified in which high or unclear risk of bias might cause shrinkage of observed treatment effects. These sources of bias are reflected by high RORs.

In this study, diminished treatment effects by relative 13% were associated with high or unclear risk of bias from reporting of incomplete outcome data. This characteristic was judged as possibly associated with high risk of bias if published study reports were restricted to results of all treated or all eligible patients or reporting was restricted to lethality or indication-specific mortality. Furthermore, an underestimation may be associated with allocation concealment.

Little evidence of biased effect estimates on all-cause mortality was found by the five characteristics: sequence generation, double blinding, early stopping or inadequate reporting of sample size, competing interests, and insufficient length of follow-ups.

A high degree of confounding through limitations in single-center trials in nearly all characteristics and between characteristics allocation concealment, sequence generation, double blinding, and estimated treatment may cause different estimates of ROR in the model with one single and 12 characteristics. These three characteristics are frequently reported with similar risk of bias. In addition, allocation concealment and sequence generation are regularly reported in consecutive sentences of the methods section of an article and risk of bias judgments on most articles show concordant results. At least someway, judgments may result from an association with subsequent flaws in conduct and reporting of the trial. These flaws might have additionally led to missing reporting of other characteristics as preplanned selective outcome reporting, sample size, incomplete baseline tables, or insufficient lengths of follow-up periods and cause confounding between risks of bias in these characteristics and treatment effect estimates.

However, results of this study suggest that incorporation of risk of bias into the interpretation of estimated treatment effect on all-cause mortality is essential.

4.1. Results in context with other meta-epidemiological studies

Several meta-epidemiological approaches concerning information on risk of bias have been proposed, but until now, they are not recommended for standard use. All the investigated associations are observational and suffer from bias of confounding between highly correlated trial characteristics [45].

Meta-epidemiological studies have provided evidence that high risk of bias in some of the investigated characteristics is associated with overoptimistic estimates in the effect of interventions assessed in clinical trials [7,19,21,23,25,46].

However, the evidence is not consistent for all trial characteristics. Bias varies between meta-analyses according to the nature of the selected interventions, outcome, and definitions of judgment on characteristics.

Our results are consistent to two previous studies [8,18]. Bellomo et al. [8] gave a lot of causes and examples from the intensive care medicine in which large treatment effects from single-center studies have been contradicted in other settings. This was confirmed by Dechartres et al. [18] in a meta-epidemiological study of therapeutic or preventive interventions on the basis of 421 RCTs in a broad medical area. In contrast, Savovic et al. [46] found in a summary from seven meta-epidemiological studies no conclusive difference of treatment effect estimates on all-cause mortality between single-center and multicenter trials (ROR, 0.90; 95% CI: 0.69, 1.17).

Despite a general association between allocation concealment and treatment effect, the effect on all-cause mortality seems to be independent of inadequate or unclear allocation concealment [7,11,46]. Our results are consistent with these results from a wide range of interventions.

Trial characteristic early stopping of a trial has been removed from the item ‘‘other risk of bias’’ of the newest Cochrane Collaboration’s tool for assessing risk of bias. Simulation evidence suggests that inclusion of early stopped trials in meta-analyses will not lead to substantial bias and exclusion of early stopped trials has the potential to bias meta-analyses toward the null and may lead to loss of precision [14].

Generally, confounding between trial characteristics might occur because studies at high risk of bias due to a particular flaw in their conduct are more likely than others to have flaws in other characteristics. This possibility was addressed by Wood et al. [7] who found a moderate confounding between inadequate allocation concealment and blinding and concluded that inadequate allocations concealment might still be considered as a proxy for studies at higher risk of bias. Schulz et al. [47] argued that failure to report exclusions in trials may have been a marker of poor trial conduct rather than true absence of exclusions and will be frequently confounded with other characteristics.

4.2. Strength and weaknesses of this study

We frequently rated trial characteristics as ‘‘unclear’’ risk of bias. This may reflect the nature of these characteristics or the insufficient reporting of design, conduct, analyses, and findings. Reported quality of a trial may not reflect the way that the trial was conducted in practice [48]. Well-conducted trials may be reported badly, and more detailed information can be received by authors of included trials to reduce judgments of unclear risk of bias.

All presented statistical models assume that all trials are fully characterized by cross tables representing treatment and event counts, and a recalculation of individual patient data is possible. Only RCTs with binary outcome measures conform to this format.

As biases in different characteristics often co-occur, suspected biases in characteristics were evaluated separately and simultaneously. Because of the restricted number of trials and meta-analyses per review, the model was restricted to three sampling levels (meta-analyses, trial, and individual patients). An analysis of interactions between different characteristics was impossible because of the high number of characteristics and possible interactions. All included reviews were published in The Cochrane Library, a similar methodology as described in the Cochrane handbook [14] can be assumed, and trials of all meta-analyses were judged independent from their review question. Therefore, an analysis of interactions between judgments of characteristics and meta-analysis was abandoned. Interactions between treatment and meta-analyses were included to model different treatment effects depending on investigated interventions.

The random-effects model incorporates between-study heterogeneity and considers an individual baseline risk of bias depending on study design, characteristics of the intervention, definition of inclusion and exclusion criteria,

comparison group, and different lengths of follow-up periods for outcome measurement of each trial. The heterogeneity between meta-analyses was reduced by the inclusion of meta-analyses from a small area of medicine with the same outcome measures and similar procedures in meta-analyses as recommended by Welton et al. [49]. The author has presented a Bayesian hierarchical meta-analytic model with random effects for a combined analysis of evidence from RCTs categorized as being at either low or high risk of bias. These models base on evidence of previously published meta-epidemiological work and assume a similar mean bias in the investigated clinical area. This strong assumption was not assured in this study.

Finally, inference on the basis of between-study differences is potentially misleading because studies differ in lots of other factors that could explain the interaction. This is a question that can be better investigated in clinical trials or a systematic review on individual patient data in which heterogeneity of patients in single-center and multicenter trials can be compared.

4.3. Implications for practice

In daily clinical practice, treatment of sepsis in intensive care unit is characterized by a broad approach of therapeutic measures and interventions, which are nowadays commonly based on early goal treatment (EGDT) considerations. During the three decades, several pathophysiological concepts had been introduced into clinical practice and consecutively a conflicting body of evidence has been generated. Those consecutive RCTs regarding therapeutic approaches from EGDT over immunoglobulins up to blood glucose control had been accompanied by controversial study results over the last years. For this reason, it seems to be sensible to address the issue of methodological problems in clinical trials conducted in the context of intensive care medicine. In this special field of clinical investigation, it seems to be much more reasonable and necessary than in other clinical fields (i.e., cardiology, oncology) that small investigator-initiated trials (IIT) are conducted for hypothesis generation. But because of methodological and also economic reasons, larger multicenter RCTs in the field of intensive care medicine are less likely to be successfully conducted. Especially, this last point may raise the question, whether these circumstances facilitate a trend of positive outcome reporting in smaller IITs having in mind that a larger multicenter trial with an anticipated negative result has a very low probability to be founded. This very considerable phenomenon could at least be demonstrated by the IABP-SHOCK (Prondzinsky et al. [50]) and IABP-SHOCK-Trial (Thiele et al. [51]).

5. Conclusion

The results of this study provide some evidence on a substantial difference in treatment effect between single-center and multicenter trials. Single-center trials tend to provide larger treatment effect than multicenter trials. They are valuable hypothesis-generating investigations but need to be read critically and their results should be cautiously used in decision making. In fact, bias in the investigated trial characteristics may cause clinically important differences in treatment effect, which might lead to changing treatment decisions.

Appendix

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2013.05.016.

References

[1] Chalmers I. Comparing like with like: some historical milestones in the evolution of methods to create unbiased comparison groups in therapeutic experiments. Int J Epidemiol 2001;30:1156−64.
[2] Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332.
[3] West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, et al. Systems to rate the strength of scientific evidence. Evidence reports/technology assessments, no. 47. (Prepared by the Research Triangle Institute-University of North Carolina Evidence-based Practice Center under Contract No. 290-97-0011). Rockville, MD: Agency for Healthcare Research and Quality (AHRQ); April 2002. Available at http://www.ncbi.nlm.nih.gov/books/NBK33881/.
[4] Gluud LL. Bias in clinical intervention research. Am J Epidemiol 2006;163:493−501.
[5] Verhagen AP, de Vet HC, de Bie RA, Boers M, van den Brandt PA. The art of quality assessment of RCTs included in systematic reviews. J Clin Epidemiol 2001;54:651−4.
[6] Oxman AD, Guyatt GH. Guidelines for reading literature reviews. CMAJ 1988;138(8):697−703.
[7] Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008;336:601−5.
[8] Bellomo R, Warrillow SJ, Reade MC. Why we should be wary of single-center trials. Crit Care Med 2009;37(12):3114−9.
[9] Annane D, Bellissant E, Cavaillon JM. Septic shock. Lancet 2005;365:63−78.
[10] Cochrane database of systematic reviews. Chichester, UK: John Wiley & Sons, Ltd; 2011.
[11] Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA 2002;287:2973−82.
[12] Latronico N, Botteri M, Minelli C, Zanotti C, Bertolini G, Candiani A. Quality of reporting of randomised controlled trials in the intensive care literature. A systematic analysis of papers published in Intensive Care Medicine over 26 years. Intensive Care Med 2002;28(9):1316−23.
[13] van Nieuwenhoven CA, Buskens E, van Tiel FH, Bonten MJ. Relationship between methodological trial quality and the effects of selective digestive decontamination on pneumonia and mortality in critically ill patients. JAMA 2001;286:335−40.
[14] Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. Oxford, UK: The Cochrane Collaboration; 2011. Available at http://www.cochrane-handbook.org/.

[15] Als-Nielsen B, Gluud LL, Gluud C. [O72] Methodological quality and treatment effects in randomised trials: a review of six empirical studies Cochrane Colloquium Abstracts Journal 2004. Ottawa, Canada: 12th Cochrane Colloquium; 2004.
[16] Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. JAMA 2003;289:454−65.
[17] Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 2004;291:2457−65.
[18] Dechartres A, Boutron I, Trinquart L, Charles P, Ravaud P. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. Ann Intern Med 2011;155: 39−51.
[19] Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 2003;7:1−76.
[20] Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;282:1054−60.
[21] Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 2001;135:982−9.
[22] Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ 2003;326:1167−70.
[23] Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998;352: 609−13.
[24] Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. JAMA 2005;294:2203−9.
[25] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273: 408−12.
[26] Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. Int J Epidemiol 2005;34:79−87.
[27] Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, Krebs Seida J, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. BMJ 2009;339:b4012.
[28] Fergusson D, Aaron SD, Guyatt G, Hebert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. BMJ 2002;325:652−4.
[29] Rosen M. The aprotinin saga and the risks of conducting meta-analyses on small randomised controlled trials—a critique of a Cochrane review. BMC Health Serv Res 2009;9:34.
[30] Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol 2000;53:1119−29.
[31] Siersma V, Als-Nielsen B, Chen W, Hilden J, Gluud LL, Gluud C. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. Stat Med 2007;26:2745−58.
[32] Kuss O. How to use SAS for logistic regression with correlated data. Proc 27th Annu SAS Users Group Int Conf (SUGI 27) 2002: 261−327.
[33] The Cochrane Collaboration. Review Manager (RevMan) [Computer program]. Version 5.1. Copenhagen, Denmark: The Nordic Cochrane Centre; 2012. Available at http://ims.cochrane.org/revman.
[34] IBM SPSS Statistics. Ehningen: IBM Deutschland GmbH 2011. Available at http://www-01.ibm.com/software/de/analytics/spss/products/statistics/.
[35] SAS version 9.2. Cary, NC: SAS Institute; 2011. Available at http://www.sas.com/.
[36] Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557−60.
[37] Alejandria MM, Lansang MA, Dans LF, Mantaring JB. Intravenous immunoglobulin for treating sepsis and septic shock. Cochrane Database Syst Rev 2002;1:CD001090.
[38] Annane D, Bellissant E, Bollaert PE, Briegel J, Keh D, Kupfer Y. Corticosteroids for treating severe sepsis and septic shock. Cochrane Database Syst Rev 2004;1:CD002243.
[39] Boeuf B, Poirier V, Gauvin F, Guerguerian AM, Roy C, Farrell CA, et al. Naloxone for shock. Cochrane Database Syst Rev 2003;4: CD004443.
[40] Havel C, Arrich J, Losert H, Gamper G, Mullner M, Herkner H. Vasopressors for hypotensive shock. Cochrane Database Syst Rev 2011;5:CD003709.
[41] Marti-Carvajal A, Salanti G, Cardona AF. Human recombinant activated protein C for severe sepsis. Cochrane Database Syst Rev 2008;1:CD004388.
[42] Unverzagt S, Machemer MT, Solms A, Thiele H, Burkhoff D, Seyfarth M, et al. Intra-aortic balloon pump counterpulsation (IABP) for myocardial infarction complicated by cardiogenic shock. Cochrane Database Syst Rev 2011;7:CD007398.
[43] Higgins JPT, Altman DG, Sterne JAC. Chapter 8: assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. Oxford, UK: The Cochrane Collaboration; 2011. Available at http://handbook.cochrane.org/chapter_8/8_assessing_risk_of_bias_in_included_studies.htm.
[44] von Elm E, Rollin A, Blumle A, Huwiler K, Witschi M, Egger M. Publication and non-publication of clinical trials: longitudinal study of applications submitted to a research ethics committee. Swiss Med Wkly 2008;138(13−14):197−203.
[45] Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002;21:1559−73.
[46] Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med 2012;157:429−38.
[47] Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. BMJ 1996;312:742−4.
[48] Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, et al. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. BMJ 2004;328:22−4.
[49] Welton NJ, Ades AE, Carlin JB, Altman DG, Sterne JAC. Models for potentially biased evidence in meta-analysis using empirically based priors. J R Stat Soc Ser A 2009;172(1):119−36.
[50] Prondzinsky R, Lemm H, Swyter M, Wegener N, Unverzagt S, Carter JM, et al. Intra-aortic balloon counterpulsation in patients with acute myocardial infarction complicated by cardiogenic shock: the prospective, randomized IABP SHOCK Trial for attenuation of multiorgan dysfunction syndrome. Crit Care Med 2010;38(1): 152−60.
[51] Thiele H, Zeymer U, Neumann FJ, Ferenc M, Olbrich HG, Hausleiter J, et al. Intraaortic balloon support for myocardial infarction with cardiogenic shock. N Engl J Med 2012;367:1287−96.

Unverzagt S, Oemler M, Braun K, Klement A (2014) Strategies for guideline implementation in primary care focusing on patients with cardiovascular disease: a systematic review. Fam Pract 31: 247–66.

Abstract

Background. Guidelines should reduce inappropriate practice and improve the efficiency of treatment. Not only methodological quality but also acceptance and successful implementation in daily practice are crucial for the benefit on patients. Focusing on cardiovascular diseases (CVD), it is still unclear which implementation strategy can improve physician adherence to the recommendations of guidelines in primary care.

Methods. We conducted a systematic review on randomized controlled trials about guideline implementation strategies on CVD. Medline, Embase, CENTRAL, conference proceedings and registers of ongoing studies were searched.

Results. Eighty-four trials met our predefined inclusion criteria, of them 54 trials compared unimodal strategies and 30 multimodal strategies to usual care. Concerning unimodal strategies, 15 trials investigated provider reminder systems, 3 audit and feedback, 15 provider education, 4 patient education, 5 promotion of self-management and 14 organizational change. The strongest benefit of a unimodal implementation strategy was found due to organizational change (odds ratio 1.96; 95% CI 1.4 to 2.75), followed by patient education, provider education and provider reminder systems. Trials on the efficacy of audit and feedback and patient self-management showed differing results or small advantages in terms of physician adherence. Multimodal interventions showed almost similar effect measures and ranking of strategies.

Conclusion. The use of implementation strategies for the distribution of guidelines on CVD can be convincingly effective on physician adherence, regardless whether based on a unimodal or multimodal design. Three distinct strategies should be well considered in such an attempt: organizational changes in the primary care team, patient education and provider education.

Key words: Cardiovascular disease, guidelines, meta-analysis, physician adherence, primary care, systematic review.

Unverzagt S, Peinemann F, Oemler M, Braun K, Klement A (2014) Meta-regression analyses to explain statistical heterogeneity in a systematic review of strategies for guideline implementation in primary care. Plos One 9: e110619.

OPEN ACCESS Freely available online

PLOS ONE

Meta-Regression Analyses to Explain Statistical Heterogeneity in a Systematic Review of Strategies for Guideline Implementation in Primary Health Care

Susanne Unverzagt[1]*, Frank Peinemann[2], Matthias Oemler[3], Kristin Braun[3], Andreas Klement[3]

1 Institute of Medical Epidemiology, Biostatistics and Informatics, University Halle/Wittenberg, Halle (Saale), Germany, **2** Children's Hospital, University of Cologne, Cologne, Germany, **3** Section of General Practice, Institute of Medical Epidemiology, Biostatistics and Informatics, University Halle/Wittenberg, Halle (Saale), Germany

Abstract

This study is an in-depth-analysis to explain statistical heterogeneity in a systematic review of implementation strategies to improve guideline adherence of primary care physicians in the treatment of patients with cardiovascular diseases. The systematic review included randomized controlled trials from a systematic search in MEDLINE, EMBASE, CENTRAL, conference proceedings and registers of ongoing studies. Implementation strategies were shown to be effective with substantial heterogeneity of treatment effects across all investigated strategies. Primary aim of this study was to explain different effects of eligible trials and to identify methodological and clinical effect modifiers. Random effects meta-regression models were used to simultaneously assess the influence of multimodal implementation strategies and effect modifiers on physician adherence. Effect modifiers included the staff responsible for implementation, level of prevention and definition pf the primary outcome, unit of randomization, duration of follow-up and risk of bias. Six clinical and methodological factors were investigated as potential effect modifiers of the efficacy of different implementation strategies on guideline adherence in primary care practices on the basis of information from 75 eligible trials. Five effect modifiers were able to explain a substantial amount of statistical heterogeneity. Physician adherence was improved by 62% (95% confidence interval (95% CI) 29 to 104%) or 29% (95% CI 5 to 60%) in trials where other non-medical professionals or nurses were included in the implementation process. Improvement of physician adherence was more successful in primary and secondary prevention of cardiovascular diseases by around 30% (30%; 95% CI -2 to 71% and 31%; 95% CI 9 to 57%, respectively) compared to tertiary prevention. This study aimed to identify effect modifiers of implementation strategies on physician adherence. Especially the cooperation of different health professionals in primary care practices might increase efficacy and guideline implementation seems to be more difficult in tertiary prevention of cardiovascular diseases.

Citation: Unverzagt S, Peinemann F, Oemler M, Braun K, Klement A (2014) Meta-Regression Analyses to Explain Statistical Heterogeneity in a Systematic Review of Strategies for Guideline Implementation in Primary Health Care. PLoS ONE 9(10): e110619. doi:10.1371/journal.pone.0110619

Editor: Harald Mischak, University of Glasgow, United Kingdom

Received June 19, 2014; **Accepted** September 15, 2014; **Published** October 24, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

* Email: susanne.unverzagt@medizin.uni-halle.de

Introduction

Cardiovascular diseases (CVDs) are major causes of morbidity and mortality across developed countries [1]. Various guidelines are published that provide information on health care in the prevention and treatment of CVD [2–5]. A time gap between publication and implementation in clinical practice may evolve. In a recent systematic review, we found that using implementation strategies (versus not using implementation strategies) improved the adherence of physicians to guidelines on primary care of patients with CVDs [6]. We also found a considerable variability of treatment effects among the two evaluated groups.

This heterogeneity may be an artifact of methodological factors [7–9]. These factors may include differences in study design features such as predefined primary outcomes, unit of randomization [10], duration of follow-up [11], and risk of bias. High risk of bias may change the magnitude and even the direction of treatment effects [12–16]. On the other hand, heterogenic effects of implementation strategies may result from a wide range of clinical factors such as variations in the patients with CVD in primary care, types and timing in the measurement of adherence, and special characteristics of investigated strategies [7–9,17,18]. Implementation strategies may be characterized as unimodal or multimodal or by their distinct quality improvement components (e.g. Shojania 2004 [19]). Strategies to improve physician adherence are complex interventions and an understanding of the effectiveness of these interventions is based on the assessment of what works best in different populations, circumstances and contexts [9,20,21]. These conditions may cause variations in treatment effects across different studies included in the review and provide an opportunity to identify clinical factors that may modify the treatment effect on physician adherence and increase scientific understanding [7]. Discordance between recommended and observed behaviour of physicians in the treatment of cardiovascular diseases is influenced by modifiable context-specific barriers as providers' and patients' knowledge of and attitudes towards adherence to health care recommendations and by external factors

related to the health system including lack of policy support for chronic care and prevention or limited access to health-care resources [22].

Primary care frequently occurs in small health care organizations with a team of professionals consisting of physicians, nurses, and other professionals in the primary, secondary, and tertiary prevention of CVD with a variety of symptoms and other diseases. The differences among professionals and/or professions involved in the care of patients with different indications and severity of CVD may influence the quality of dissemination and implementation of guidelines [23].

Statistical heterogeneity describes the variability in treatment effect estimates between studies and may arise from methodological or clinical heterogeneity, from other unknown, unrecorded or unreported study characteristics, or may be due to chance [17]. The causes and the extent of heterogeneity should be evaluated as they may compromise the implications of systematic reviews [24]. Statistical heterogeneity between treatment effects estimated by individual studies can be visually assessed in forest plots, tested for statistical significance, and quantified using the percentage of total variation across studies I^2 [25,26], or the between-study variance τ^2 [27,28]. Heterogeneity may be explored by conducting subgroup analyses or weighted meta-regression in complex interventions [29,30]. Meta-regression analysis can be used for a simultaneous exploration of potential methodological and clinical effect modifiers. One way of dealing with statistical heterogeneity in meta-analyses and meta-regression analyses is to incorporate a term to account for it in a random-effects model.

This study is a workup of a previously published systematic review [6]. The main aim is to explain the heterogeneity of findings from the original review to identify possible methodological and clinical heterogeneity factors that may have influenced estimated treatment effects of guideline implementation strategies on physician adherence in primary care of patients with CVD.

Methods

Study design

The design of this systematic review and the efficacy results of different implementation strategies in comparison to usual care have been previously published [6]. In brief, searches included Medline, Embase, the Cochrane Library, references of included studies, conference proceedings, register of ongoing trials and references of all included studies published between 1990 and 2012. The review considered all randomized studies that investigated guideline dissemination and their implementation into treatment of patients with CVD in primary care practices. Trials had to report guideline adherence of physicians over a minimum period of 3 months after initiating the implementation strategies. We extracted information on design, indication of patients according to the International Classification of Diseases-10 (ICD-10), implementation strategies and outcomes from all eligible trials into standardized data extraction tables. Implementation strategies were categorized as provider reminder systems, provider education, facilitated relay to data, audit and feedback, promotion of self-management, patient education, patient reminder, and organizational change [31]. The internal validity of trials was judged according to the Cochrane Collaboration risk of bias tool [10] with extension to cluster randomized trials (c-RCTs) [32–34] as high, unclear or low in six specific domains including bias in random sequence generation, allocation concealment, blinded outcome assessment, documentation of incomplete outcome data and selective reporting. Furthermore, baseline comparability between treatment groups and the use of adjustment methods to cope with potential imbalances in both cluster and individual characteristics were summarized as other sources of bias. All steps were done by two independent authors, disagreements were resolved by discussion until consensus was obtained. All treatment effect measures for the primary endpoint (physician adherence) were presented as odds ratios (OR) with 95% confidence intervals (CI) and recoded so that an OR higher than 1 indicates a beneficial effect with higher physician adherence in the experimental implementation strategy. Multiple endpoints were summarized using the mean of logarithmic ORs and ORs were recalculated from relative risks [10], standardized mean difference with standard deviation [35], and absolute frequencies of physician adherence in different groups. Results of c-RCTs without hierarchical modelling were corrected with the reported intracluster correlation coefficient and the mean number of patients per cluster [10]. Effect sizes were interpreted in categories of *small* to describe effect sizes $\leq$20%, *moderate* to describe effect sizes$>$20 and $\leq$50%, and *large* to describe effect sizes of $>$50% increase of physician adherence in comparison to passive implementation.

Investigation of heterogeneity

In our first publication we explored treatment effects in subgroups of unimodal interventions, graphically displayed them in forest plots, and quantified the remaining heterogeneity of treatment effects using the I^2 and τ^2 values [6]. Due to the high heterogeneity between included trials we used the random-effects model for meta-analysis of the relevant comparisons of implementation strategies to usual care. This study adds a simultaneous assessment of the influence of multimodal implementation strategies and different effect modifiers on physician adherence in six random effects meta-regression models. We separately added six single sources of heterogeneity to binary variables describing the components of multimodal implementation strategies and investigated their influence on the treatment effect. Sources of heterogeneity included the staff responsible for implementation, level of prevention, definition of the primary outcome, unit of randomization, duration of follow-up and risk of bias domains.

We quantified the influence of all investigated sources of heterogeneity with ratios of odds ratios (ROR) comparing OR of studies with different values of the sources of heterogeneity (e.g. studies in secondary prevention with those in tertiary prevention of CVD). ROR with 95% confidence intervals (95%CI) not containing the null value (ROR = 1) will be interpreted as significant. Heterogeneity was measured using the τ^2 statistics which estimates the between-trial variability. The amount of heterogeneity explained by different effect modifiers was described by the relative reduction of τ^2. All modifiers were investigated as binary traits. Categorical traits (staff responsible for implementation, type of prevention) were recoded into dummy variables. We conducted the statistical analyses using RevMan5 for the systematic review and SAS 9.2 (PROC MIXED statement) for this study.

Results

Meta-analyses of treatment effects on physician adherence revealed considerable heterogeneity in all subgroups of unimodal interventions [6]. In this study, altogether 75 trials were pooled in 84 comparisons between unimodal or multimodal active and passive implementation strategies (Table S1). Of these comparisons, 13 indicated a negative treatment effect with an OR$<$1, 17 a small effect size, 22 a moderate effect size, and 32 a large effect size of physician adherence to guidelines.

In the majority of trials the implementation strategies were directed at physicians (70 trials). Physicians were supported by nurses in 19 trials or other non-medical professionals such as pharmacists (12 trials), (study)-assistants (3 trials), health workers (2 trials), (peer-) supervisors, praxis managers, or other specialists (1 trial). In four trials, strategies were exclusively implemented by specialized nurses and in one trial by a team of nurses, Asian link (health-) workers, and community diabetes specialists. Most trials concentrated on patients in the secondary prevention of CVD, 10 trials additionally included patients in primary prevention, and 6 trials included patients in tertiary prevention. Only 6 trials were limited to primary prevention and 12 trials to tertiary prevention. The most successful strategies based on patient education and organizational change regularly included nurses (46 and 41% of trials) or other professionals (31 and 71% of trials) in the implementation process (Table 1).

Approximately half of the trials were pre-planned with physician adherence as the primary outcome (36 trials) and the other half of trials reported adherence as the secondary outcome or parameter describing the process of care (39 trials). Units of analyses were individual patients in 23 RCTs and practices in 52 c-RCTs. Implementation strategies that were directed to the staff of general practices, such as provider reminder systems, audit and feedback, provider education, and organizational change, were frequently investigated in c-RCTs with physician adherence as the primary outcome. On the other hand, trials on strategies directed to patients such as patient education, promotion of self-management, and patient reminders predominantly concentrated on patient-related outcomes and reported physician adherence to the implementation process mostly as a secondary outcome. Follow-up periods between 3 and 36 months were used to investigate the efficacy of the intervention with a median length of follow-up of 12 months. Of the 75 trials, 61 trials had follow-up periods of between 3 and 12 months and 14 had longer follow-up periods of up to 36 months (Table 1).

Of the 75 trials, 48 reported the method of randomization in the text. The treatment allocation of clusters or patients was described as concealed in 63 trials. Physician adherence was assessed on objective criteria (such as number of medications, or by external monitors) and/or blinded in 63 trials. In 61 trials, the analyses were done by intention-to-treat, both at the individual and at the cluster levels. Total numbers of dropouts were low (<10%) and their causes were given by group. Primary endpoints were pre-specified in sample-size calculations and were adequately reported in 47 trials. Other sources of bias were evident in 28 trials that made no use of appropriate methods for the adjustment of treatment effects on physician adherence to cope with potential imbalances in cluster and individual characteristics. Summarizing these results, only 15 trials had low risk of bias in all the investigated domains.

Association between heterogeneity factors and estimated treatment effects

We calculated relative frequencies of negative, small, moderate, and large treatment effects depending on subgroups with special clinical and methodological characteristics. These characteristics were correlated with special implementation strategies, as shown in Table 2.

In general, the inclusion of other non-medical professionals seems to be most successful in improving physician adherence to guidelines. Large treatment effects are indicated in 57% of comparisons where non-medical professionals were included in the implementation process, compared to 38% if exclusively physicians were included, and 42% if nurses were included. Furthermore, large treatment effects are more frequently found in trials on

Table 1. Investigated sources of heterogeneity in 75 trials.

Implementation strategies			Direction of implementation	Level of prevention	Design	Investigation of physician adherence as	Follow-up
Categories	**Data source**	**Odds ratio; 95% CI**	**Physician/nurse/other professions**	**Primary/Secondary/ Tertiary**	**RCT/c-RCT**	**Primary/secondary outcome**	**3–12 months/>12 months**
Provider reminder systems	[49–70]	1.07; 0.93 to 1.23	22 (100%)/7 (32%)/5 (23%)	4 (18%)/18 (82%)/7 (32%)	3 (14%)/19 (86%)	17 (78%)/5 (22%)	20 (91%)/2 (9%)
Facilitated relay to data	[71]	2.01; 1.02–3.96	1 (100%)/0 (0%)/0 (0%)	0 (0%)/1 (100%)/0 (0%)	0 (0%)/1 (100%)	0 (0%)/1 (100%)	1 (100%)/0 (0%)
Audit and feedback	[55,56,59,63,72–79]	1.01; 0.73 to 1.40	12 (100%)/1 (8%)/3 (25%)	6 (50%)/9 (76%)/1 (8%)	2 (17%)/10 (83%)	7 (58%)/5 (42%)	9 (75%)/3 (25%)
Provider education	[49,51,59,72,73,76,78–96]	1.34; 1.08 to 1.65	23 (92%)/4 (16%)/4 (14%)	5 (20%)/15 (60%)/7 (28%)	1 (4%)/24 (96%)	16 (64%)/9 (36%)	19 (76%)/6 (24%)
Patient education	[55,72,80,86,96–105]	1.48; 1.08 to 2.01	11 (79%)/6 (46%)/4 (31%)	3 (23%)/10 (77%)/2 (14%)	5 (36%)/9 (64%)	3 (21%)/11 (79%)	10 (71%)/4 (29%)
Promotion of self-management	[71,72,83,91,97,99,102, 105–109]	1.08; 0.80 to 1.45	10 (83%)/4 (33%)/1 (8%)	1 (8%)/9 (75%)/3 (25%)	7 (58%)/5 (42%)	1 (8%)/11 (92%)	12 (100%)/0 (0%)
Patient reminders	[52,97,98,101,102,104,110]	0.81; 0.51 to 1.28	7 (100%)/4 (58%)/0 (0%)	0 (0%)/5 (83%)/2 (29%)	5 (71.4%)/2 (28.6%)	1 (14%)/6 (86%)	5 (71%)/2 (28%)
Organizational change	[69,85,93,103,111–123]	1.49; 1.21 to 1.82	15 (88%)/7 (41%)/12 (71%)	5 (29%)/14 (82%)/5 (29%)	8 (47%)/9 (53%)	6 (35%)/11 (65%)	13 (76%)/4 (24%)

Abbreviations: CI: confidence interval; c-RCT: cluster randomized controlled trial; RCT: randomized controlled trial.

doi:10.1371/journal.pone.0110619.t001

Table 2. Association between trial characteristics and treatment effect in 84 comparisons.

Potential effect modifier	Negative direction of effect	Small size	Moderate effect size	Large effect size
	OR<1.0	OR≥1.0 and OR≤1.2	OR>1.2 and OR≤1.5	OR>1.5
Implementation received by physician	12 (16%)	15 (19%)	21 (27%)	29 (38%)
Implementation received by nurse	4 (17%)	4 (17%)	6 (25%)	10 (42%)
Implementation received by other professionals	1 (4%)	6 (21%)	5 (18%)	16 (57%)
Primary prevention	2 (9%)	5 (24%)	4 (19%)	10 (48%)
Secondary prevention	11 (18%)	9 (15%)	17 (27%)	25 (40%)
Tertiary prevention	3 (14%)	7 (33%)	6 (29%)	5 (24%)
Design: c-RCT	10 (17%)	15 (25%)	14 (23%)	21 (35%)
Design: RCT	3 (12%)	2 (8%)	8 (33%)	11 (46%)
Adherence as primary outcome	5 (12%)	7 (17%)	13 (32%)	16 (39%)
Adherence as secondary outcome	8 (19%)	10 (23%)	9 (21%)	16 (37%)
Follow-up <12 months	12 (18%)	12 (18%)	19 (29%)	23 (35%)
Follow-up ≥12 months	1 (6%)	5 (28%)	3 (17%)	9 (50%)

Abbreviations: RCT: randomized controlled trial; c-RCT: cluster randomized controlled trial; OR: odds ratio.
doi:10.1371/journal.pone.0110619.t002

primary or secondary prevention (48 or 40% of trials) compared to tertiary prevention (24%) of CVD.

Comparisons of trials investigating the process of implementation with physician adherence as a primary outcome less frequently showed a negative direction of effect (12% vs. 19%) and more frequently a moderate or large effect size with OR>1.2 (71% vs. 58%). Finally, RCTs were more frequently successful with a moderate or large effect size compared to c-RCTs (79 vs. 58%), and studies with longer follow-up periods of at least 12 months more frequently showed a large effect size (50% vs. 35%) compared to shorter trials.

Table 3 summarizes the ratios of odds ratios (ROR) for the influence of all investigated clinical and methodological heterogeneity factors on treatment effect. Between-trial variability and the relative reduction of between-trial variability by single effect modifiers compared to the variability of the original model describe the reduction of statistical heterogeneity.

These results showed that the treatment effect varied depending on clinical heterogeneity factors. The receiver (i.e., person/profession responsible for implementation) of the implementation seems especially to influence the efficiency of implementation strategies on physician adherence. The inclusion of the receiver of the implementation into the meta-regression reduced between-trial variability (τ^2) by 27%. Physician adherence was improved by 62% (ROR 1.62; 95% CI 1.29 to 2.04) in trials where other non-medical professionals were included in the process of implementation, and expanding the role of nurses in the curative process was

Table 3. Association of different effect modifiers on treatment effect.

Potential effect modifier	Comparisons	ROR; 95% CI	Between-trial variability (τ^2); relative reduction
No effect modifier			0.1899
Staff	Nurse as receiver vs. others	1.29; 1.05 to 1.60	0.1389; 26.9%
	Other professionals as receiver vs. others	1.62; 1.29 to 2.04	
Level of prevention	Primary prevention vs. others	1.30; 0.98 to 1.71	0.1692; 10.9%
	Secondary prevention vs. others	1.31; 1.09 to 1.57	
Unit of randomization	c-RCT vs. RCT	1.28; 1.03 to 1.60	0.1871; 1.5%
Outcome definition	Adherence as primary vs. secondary outcome	1.38; 1.12 to 1.70	0.1719; 9.5%
Duration of follow-up	Long (>12 months) vs. short follow-up periods	1.38; 1.04 to 1.83	0.1741; 8.3%
Risk of bias	Risk of bias (high or unclear vs. low):		0.1488; 21.6%
	Sequence generation	0.88; 0.77 to 1.27	
	Allocation concealment	0.93; 0.64 to 1.33	
	Blinding	1.11; 0.80 to 1.53	
	Incomplete outcome data addressed	1.04; 0.77 to 1.40	
	Selective outcome reporting	1.58; 0.96 to 2.60	
	Other	0.94; 0.74 to 1.20	

Abbreviation: CI: confidence interval; c-RCT: cluster randomized controlled trial; RCT: randomized controlled trial; ROR: ratio of odds ratios.
doi:10.1371/journal.pone.0110619.t003

successful in improving physician adherence by 29% (ROR 1.29; 95% CI 1.05 to 1.60).

Inclusion of level of prevention as a clinical effect modifier reduced τ^2 by 11%. Improvement of physician adherence was most successful in the treatment of patients in primary and secondary prevention of CVD (ROR 1.30; 95% CI 0.98 to 1.71 and 1.31; 95% CI 1.09 to 1.57) compared to patients in tertiary prevention.

However, methodological issues seem to have a smaller influence on the investigated association. We found an association between the definition of the primary endpoint as a quantitative measure of the main objective of a trial and the estimated implementation effect. The inclusion of this effect modifier was able to reduce τ^2 by 10%. The implementation effect on the primary outcome in process optimization studies with the primary endpoint of physician adherence was increased by 38% (ROR 1.38; 95% CI 1.12 to 1.70) compared to outcome optimization studies where physician adherence was investigated as a secondary or process of care parameter. Moreover, we also found an increased implementation effect in RCTs compared to c-RCTs by 28% (ROR 1.28; 95% CI 1.02 to 1.60) and with longer duration of follow-up (ROR 1.38; 95% CI 1.03 to 1.83), with only small reductions of τ^2 by 2 and 8%, respectively.

Inclusion of six risk of bias domains reduced between-trial variability by 22%, but no single component was associated with a significant overestimation of implementation effect. We found a tendency to an overestimation of implementation effects in trials with potential bias by selective reporting of the primary outcome by 58% (ROR 1.58; 95% CI 0.96 to 2.60).

Discussion

This study is based on an analysis of 75 trials and on eight classes of implementation strategies to improve physician's guideline adherence. It investigates the influence of six possible effect modifiers on estimated implementation effects and remaining statistical heterogeneity. These influences are quantified by RORs and the relative change of between trial variability. Our investigations revealed a substantial reduction of statistical heterogeneity explained by five of the investigated effect modifiers.

We found that clinical effect modifiers such as the cooperation of physicians with non-medical health professionals, the setting of the primary care of patients in early prevention of CVD, and the duration of implementation were especially associated with the improvement of physicians' adherence to guidelines. We have found a considerable reduction of statistical heterogeneity by these factors of 27, 11, and 8%, respectively. Furthermore, the inclusion of methodological effect modifiers as different sources of bias or the definition of the primary endpoint was able to reduce statistical heterogeneity by 22 and 8%, respectively.

A considerable amount of statistical heterogeneity is explained by organizational structures in the primary care practices. Improvement of adherence could be achieved if physicians accepted support from non-medical health professionals such as pharmacists, health workers, qualified nurses, or nurse practitioners in improving their professional and organizational performance. Such cooperation could take place within team or (smaller) teamlet structures in single practices or networks of care [36]. Teamwork among different professions and/or professionals aimed at implementing guidelines explains the different estimated treatment effects on the adherence of primary care physicians. Our results are in line with the conclusions of Grol and Grimshaw (2003) [37] and Unverzagt et al. (2013) [6] which suggest that the whole primary care team (or network) is important for the implementation success.

We have further showed that different levels of prevention may cause heterogeneous treatment effects with greater improvements in physician adherence in the primary care of patients in the early prevention of CVD. Guidelines are more difficult to implement in tertiary prevention, where patients frequently suffer from complex co-morbidities and guidance on interventions and information on risks of specific interventions is missed [38]. This observation reflects the problem how far results from RCT can be generalized due to selection in patients included in RCTs and narrow inclusion/exclusion criteria [39]. Especially patients with complex comorbidities should be well represented in RCTs and guidelines should provide guidance for the often complex need of these patients including information on risks of specific interventions [38].

Moreover, our explorations revealed artificial sources of variation resulting from the inclusion of two different types of study with different main aims, although both types reported physician adherence. Physician adherence summarizes the degree of conformity between knowledge, cognition, and action in a primary physician center with evidence-based recommendations from guidelines on different aspects of quality of care [40].

The primary aim of the first type of study is to improve the care process by implementing guidelines and the primary aim of the second type is to improve the health outcomes of patients. A "change" to a more evidence-based treatment aimed at improving health outcomes is a stepwise complex process in which several individual and organizational barriers have to be removed and intermediate outcomes (such as adherence) and final outcomes (such as health outcomes) may be improved. First, different strategies should be used to implement evidence-based guidelines in the care process and to enhance the adherence of physicians. Secondly, patients must be adherent to evidence-based recommendations, and finally, the health outcomes for patients might change. The benefit they receive from guideline-oriented treatment by reduced hospital admissions and prolonged survival has been shown in several studies [41–43]. In trials designed to improve the implementation of guidelines, physician adherence was regularly chosen as a primary endpoint and measured by multiple indicators, which were summarized to adherence scores or described by different frequencies.

The second type of study concentrates on the improvement of health outcomes of patients and describes physician adherence as a step towards improving mortality, morbidity, health-related quality of life, or surrogate parameters such as the achievement of targets for blood-pressure, cholesterol concentrations, physical activity, body mass index, smoking cessation or reduced smoking, or the reduction of cardiovascular risk score. Some statistical heterogeneity of treatment effects can be explained by these different types of studies.

In addition, we stated an increased treatment effect in studies with longer follow-up periods where implementation strategies were used over a longer period with a potentially better chance to improve providers knowledge, to integrate guidelines recommendation into organizational structures and processes and to overcome as well as negative staff attitudes and beliefs and time and resource constraints in primary care centers.

Furthermore, we identified an influence of risk of bias on the variability of treatment effect and were able to reduce statistical heterogeneity, but we were not able to identify one single source of bias causing biased treatment effects. However, different additional factors show divergent influences on the process of implementation and health outcomes and may influence both the process of

implementation and the efficacy of the recommended intervention for the patient. These factors include concerns about the quality of guidelines such as the quality of evidence on which they are based, lack of agreement, differences in strength of recommendations, practicability of guidelines and recommended interventions, and the benefit/harm ratio of the intervention [44,45]. Finally, financial constraints and organizational structures (e.g., health systems) may modify the process of care [46].

Strengths and limitations

We are aware that our research may have some limitations. The conclusions of our systematic review [6] were limited by a large amount of imprecision and inconsistency of reported results [24,47,48]. Unexplained heterogeneity was caused by different methodological and clinical effect modifiers and some of them were investigated in this study. Other, not investigated potential effect modifiers are differences in health care systems, stability, attitudes and resource constraints of health care teams, types of guideline-recommended interventions, patient decisions and treatment. This study contributes an improved insight into the processes and elements of successful change and shows the environment in which implementation strategies work best, but all these ideas are exploratory and should be considered as hypotheses for evaluation in future studies [17]. Developing of precise definitions of effect modifiers for complex definitions of participants, intervention and outcomes was problematic because our priori definitions were influenced from our growing knowledge on existing evidence. To reduce the dependence on the results of data syntheses, we discussed these potential effect modifiers and tried to find the best classifications on the basis of diverse extracted clinical and methodological characteristics before we started data synthesis. Moreover, we are assured that the investigated heterogeneity factors are unlikely to influence the pathway from implementation strategy to an improved physician adherence and patient outcome independently and may reflect associations with other correlated but not investigated factors [25]. Furthermore, it is possible that especially in studies with inclusion of patients in early and late prevention relationships across trials may not be the same as relationships for patients within trials [29].

Conclusions

We recommend a careful discussion of the pathway from the intervention to the outcomes and pre-definition of potential effect modifiers at an early time point when conducting a systematic review. This seems to be especially important in a review of complex interventions with a broad range of participants, interventions, comparators, and definitions of outcomes. Our work has explained some statistical heterogeneity by clinical and methodological effect modifiers. These effect modifiers cause variability in estimated treatment effects, and taking them into account reduced statistical heterogeneity.

The results of this study provide some evidence that the incorporation of different health professionals in the practice can change professional and organizational performance in a primary care practice. This study properly investigated the role of implementation strategies in the treatment of patients in primary and secondary prevention of CVD and provides some evidence of promising results. However, the implementation of guidelines into tertiary prevention of CVD in general practice requires improved guidance for this patient group. Finally, we propose choosing the process parameter of physician adherence as the primary outcome parameter only in cases where a theoretical model explains the route from the intervention to the anticipated health outcomes for patients. Nevertheless, the health outcomes for patients should always be measured and reported additionally to process parameters (e.g., physician adherence) to ensure the association between the process parameter and improved health outcomes and to improve understanding of the pathway from implementation over physician adherence to improved health outcomes.

Supporting Information

Table S1 Included studies and comparisons with treatment effect, implementation strategies in the intervention group, risk of bias and effect modifiers. Abbreviations: lnOR: logarithmic Odds Ratio, SElnOR: standard error of logarithmic Odds Ratio.
(XLSX)

Acknowledgments

We thank Roland Prondzinsky for his constructive comments and providing clinical background.

Author Contributions

Conceived and designed the experiments: SU. Performed the experiments: SU. Analyzed the data: SU. Contributed reagents/materials/analysis tools: SU. Wrote the paper: SU FP MO KB AK.

References

1. Murray CJL, Lopez AD (1996) Evidence-based health policy - lessons from the global burden of disease study. Science 274: 740–743.
2. AHA (2014) ACC/AHA Joint Guidelines. Dallas: American Heart Association.
3. CDC (2014) Prevention. Guidelines and Recommendations. Atlanta: Centers for Disease Control and Prevention.
4. ESC (2014) ESC Clinical Practice Guidelines section. Sophia Antipolis: European Society of Cardiology.
5. WHO (2014) Prevention of cardiovascular disease (CVDs). Geneva: World Health Organization.
6. Unverzagt S, Oemler M, Braun K, Klement A (2013) Strategies for guideline implementation in primary care focusing on patients with cardiovascular disease: a systematic review. Fam Pract 31: 247–266.
7. Thompson SG (1994) Why sources of heterogeneity in meta-analysis should be investigated. BMJ 309: 1351–1355.
8. Glasziou PP, Sanders SL (2002) Investigating causes of heterogeneity in systematic reviews. Stat Med 21: 1503–1511.
9. Pigott T, Shepperd S (2013) Identifying, documenting, and examining heterogeneity in systematic reviews of complex interventions. J Clin Epidemiol 66: 1244–1250.
10. Cochrane Collaboration (2011) Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]; Higgins JPT, Green S, editors. Oxford: The Cochrane Collaboration.
11. Rosén M (2009) The aprotinin saga and the risks of conducting meta-analyses on small randomised controlled trials - a critique of a Cochrane review. BMC Health Serv Res 9: 34.
12. Schulz KF, Chalmers I, Hayes RJ, Altman DG (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 273: 408–412.
13. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, et al. (1998) Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 352: 609–613.
14. Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, et al. (2002) Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA 287: 2973–2982.
15. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, et al. (2008) Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 336: 601–605.
16. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J (2003) How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 7: 1–76.
17. Gagnier JJ MH, Altman DG, Berlin J, Chang S, McCulloch P, et al. (2013) Consensus-based recommendations for investigating clinical heterogeneity in systematic reviews. BMC Med Res Methodol 13.

18. Higgins J, Thompson S, Deeks J, Altman D (2002) Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. J Health Serv Res Policy 7: 51–61.
19. Shojania KG, McDonald KM, Wachter RM, Owens DK (2004) Closing Closing The Quality Gap: A Critical Analysis of Quality Improvement Strategies, Volume 1—Series Overview and Methodology. Technical Review 9 (Contract No. 290-02-0017 to the Stanford University–UCSF Evidence-based Practices Center). AHRQ Publication No. 04-0051-1. Rockville: Agency for Healthcare Research and Quality.
20. Petticrew M, Anderson L, Elder R, Grimshaw J, Hopkins D, et al. (2013) Complex interventions and their implications for systematic reviews: a pragmatic approach. J Clin Epidemiol 66: 1209–1214.
21. Petticrew M, Rehfuess E, Noyes J, Higgins JP, Mayhew A, et al. (2013) Synthesizing evidence on complex interventions: how meta-analytical, qualitative, and mixed-method approaches can contribute. J Clin Epidemiol 66: 1230–1243.
22. Nieuwlaat R, Schwalm JD, Khatib R, Yusuf S (2013) Why are we failing to implement effective therapies in cardiovascular disease? Eur Heart J 34: 1262–1269.
23. Wensing M, van der Weijden T, Grol R (1998) Implementing guidelines and innovations in general practice: which interventions are effective? Br J Gen Pract 48: 991–997.
24. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, et al. (2011) GRADE guidelines: 7. Rating the quality of evidence–inconsistency. J Clin Epidemiol 64: 1294–1302.
25. Higgins JP, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. Stat Med 21: 1539–1558.
26. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327: 557–560.
27. DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. Control Clin Trials 7: 177–188.
28. Rucker G, Schwarzer G, Carpenter JR, Schumacher M (2008) Undue reliance on I(2) in assessing heterogeneity may mislead. BMC Med Res Methodol 8: 79.
29. Thompson SG, Higgins JP (2002) How should meta-regression analyses be undertaken and interpreted? Stat Med 21: 1559–1573.
30. Bower P, Gilbody S, Richards D, Fletcher J, Sutton A (2006) Collaborative care for depression in primary care. Making sense of a complex intervention: systematic review and meta-regression. Br J Psychiatry 189: 484–493.
31. AHRQ (2004) Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies (Vol. 1: Series Overview and Methodology); Shojania KG, McDonald KM, Wachter RM, Owens DK, editors. Rockville: Agency for Healthcare Research and Quality.
32. Puffer S, Torgerson D, Watson J (2003) Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. BMJ 327: 785–789.
33. Giraudeau B, Ravaud P (2009) Preventing bias in cluster randomised trials. PLoS Med 6: e1000065.
34. Campbell MK, Piaggio G, Elbourne DR, Altman DG, Group C (2012) Consort 2010 statement: extension to cluster randomised trials. BMJ 345: e5661.
35. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009) Introduction to meta-analysis. Chichester: John Wiley & Sons.
36. Chen EH, Thom DH, Hessler DM, Phengrasamy L, Hammer H, et al. (2010) Using the Teamlet Model to Improve Chronic Care in an Academic Primary Care Practice. J Gen Intern Med 25: 610–614.
37. Grol R, Grimshaw J (2003) From best evidence to best practice: effective implementation of change in patients' care. Lancet 362: 1225–1230.
38. Francke AL, Smit MC, de Veer AJ, Mistiaen P (2008) Factors influencing the implementation of clinical guidelines for health care professionals: a systematic meta-review. BMC Med Inform Decis Mak 8: 38.
39. Bailey KR (1994) Generalizing the results of randomized clinical trials. Control Clin Trials 15: 15–23.
40. Hasenbein U, Wallesch CW (2007) [What is "adherence to guidelines"? Theoretical and methodological considerations on a new concept for health system research and quality management]. Gesundheitswesen 69: 427–437.
41. Kuepper-Nybelen J, Hellmich M, Abbas S, Ihle P, Griebenow R, et al. (2012) Association of long-term adherence to evidence-based combination drug therapy after acute myocardial infarction with all-cause mortality. A prospective cohort study based on claims data. Eur J Clin Pharmacol 68: 1451–1460.
42. Frankenstein L, Remppis A, Fluegel A, Doesch A, Katus HA, et al. (2010) The association between long-term longitudinal trends in guideline adherence and mortality in relation to age and sex. Eur J Heart Fail 12: 574–580.
43. Komajda M, Lapuerta P, Hermans N, Gonzalez-Juanatey JR, van Veldhuisen DJ, et al. (2005) Adherence to guidelines is a predictor of outcome in chronic heart failure: the MAHLER survey. Eur Heart J 26: 1653–1659.
44. Ploeg J, Davies B, Edwards N, Gifford W, Miller PE (2007) Factors influencing best-practice guideline implementation: lessons learned from administrators, nursing staff, and project leaders. Worldviews Evid Based Nurs 4: 210–219.
45. Prior M, Guerin M, Grimmer-Somers K (2008) The effectiveness of clinical guideline implementation strategies–a synthesis of systematic review findings. J Eval Clin Pract 14: 888–897.
46. SIGN (2008) SIGN 50: A guideline developer's handbook. Edinburgh: Scottish Intercollegiate Guidelines Network.
47. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, et al. (2011) GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 64: 401–406.
48. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, et al. (2011) GRADE guidelines 6. Rating the quality of evidence–imprecision. J Clin Epidemiol 64: 1283–1293.
49. Beaulieu MD, Brophy J, Jacques A, Blais R, Battista R, et al. (2004) Drug treatment of stable angina pectoris and mass dissemination of therapeutic guidelines: a randomized controlled trial. QJM 97: 21–31.
50. Bebb C, Kendrick D, Coupland C, Madeley R, Stewart J, et al. (2007) A cluster randomised controlled trial of the effect of a treatment algorithm for hypertension in patients with type 2 diabetes. Br J Gen Pract 57: 136–143.
51. Casebeer LL, Klapow JC, Centor RM, Stafford MA, Renkl LA, et al. (1999) An intervention to increase physicians' use of adherence-enhancing strategies in managing hypercholesterolemic patients. Acad Med 74: 1334–1339.
52. Cobos A, Vilaseca J, Asenjo C, Pedro-Botet J, Sanchez E, et al. (2005) Cost effectiveness of a clinical decision support system based on the recommendations of the European Society of Cardiology and other societies for the management of hypercholesterolemia: Report of a cluster-randomized trial. Dis Manag Health Out 13: 421–432.
53. Eccles M, McColl E, Steen N, Rousseau N, Grimshaw J, et al. (2002) Effect of computerised evidence based guidelines on management of asthma and angina in adults in primary care: Cluster randomised controlled trial. BMJ 325: 941–944.
54. Frances CD, Alperin P, Adler JS, Grady D (2001) Does a fixed physician reminder system improve the care of patients with coronary artery disease? A randomized controlled trial. West J Med 175: 165–166.
55. Fretheim A, Oxman AD, Havelsrud K, Treweek S, Kristoffersen DT, et al. (2006) Rational prescribing in primary care (RaPP): A cluster randomized trial of a tailored intervention. PLoS Med 3: 783–791.
56. Frijling BD, Lobo CM, Hulscher ME, Akkermans RP, van Drenth BB, et al. (2003) Intensive support to improve clinical decision making in cardiovascular care: a randomised controlled trial in general practice. Qual Saf Health Care 12: 181–187.
57. Gill JM, Chen YX, Glutting JJ, Diamond JJ, Lieberman MI (2009) Impact of decision support in electronic medical records on lipid management in primary care. Popul Health Manag 12: 221–226.
58. Gilutz H, Novack L, Shvartzman P, Zelingher J, Bonneh DY, et al. (2009) Computerized community cholesterol control (4C): Meeting the challenge of secondary prevention. Isr Med Assoc J 11: 23–29.
59. Goff DC Jr, Gu L, Cantley LK, Sheedy DJ, Cohen SJ (2003) Quality of care for secondary prevention for patients with coronary heart disease: results of the Hastening the Effective Application of Research through Technology (HEART) trial. Am Heart J 146: 1045–1051.
60. Heidenreich PA, Gholami P, Sahay A, Massie B, Goldstein MK (2007) Clinical reminders attached to echocardiography reports of patients with reduced left ventricular ejection fraction increase use of beta-blockers: A randomized trial. Circulation 115: 2829–2834.
61. Hicks LS, Sequist TD, Ayanian JZ, Shaykevich S, Fairchild DG, et al. (2008) Impact of computerized decision support on blood pressure management and control: a randomized controlled trial. J Gen Intern Med 23: 429–441.
62. Hung CS, Lin JW, Hwang JJ, Tsai RY, Li AT (2008) Using paper chart based clinical reminders to improve guideline adherence to lipid management. J Eval Clin Pract 14: 861–866.
63. Lester WT, Grant RW, Octo BG, Chueh HC (2006) Randomized controlled trial of an informatics-based intervention to increase statin prescription for secondary prevention of coronary disease. J Gen Intern Med 21: 22–29.
64. Montgomery AA, Fahey T, Peters TJ, MacIntosh C, Sharp DJ (2000) Evaluation of computer based clinical decision support system and risk chart for management of hypertension in primary care: randomised controlled trial. BMJ 320: 686–690.
65. Murray MD, Harris LE, Overhage JM, Zhou XH, Eckert GJ, et al. (2004) Failure of computerized treatment suggestions to improve health outcomes of outpatients with uncomplicated hypertension: results of a randomized controlled trial. Pharmacotherapy 24: 324–337.
66. Rossi RA, Every NR (1997) A computerized intervention to decrease the use of calcium channel blockers in hypertension. J Gen Intern Med 12: 672–678.
67. Santschi V, Rodondi N, Bugnon O, Burnier M (2008) Impact of electronic monitoring of drug adherence on blood pressure control in primary care: a cluster 12-month randomised controlled study. Eur J Intern Med 19: 427–434.
68. Sequist TD, Gandhi TK, Karson AS, Fiskio JM, Bugbee D, et al. (2005) A randomized trial of electronic clinical reminders to improve quality of care for diabetes and coronary artery disease. J Am Med Inform Assoc 12: 431–437.
69. Tierney WM, Overhage JM, Murray MD, Harris LE, Zhou XH, et al. (2003) Effects of computerized guidelines for managing heart disease in primary care. J Gen Intern Med 18: 967–976.
70. Van Wyk JT, van Wijk MAM, Sturkenboom MC, Mooseveld M, Moormann PW, et al. (2008) Electronic Alerts Versus On-Demand Decision Support to Improve Dyslipidemia Treatment: A Cluster Randomized Controlled Trial. Circulation 117: 371–378.
71. Villeneuve J, Genest J, Blais L, Vanier MC, Lamarre D, et al. (2010) A cluster randomized controlled Trial to Evaluate an Ambulatory primary care Management program for patients with dyslipidemia: the TEAM study. CMAJ 182: 447–455.

72. Benner JS, Erhardt L, Flammer M, Moller RA, Rajicic N, et al. (2008) A novel programme to evaluate and communicate 10-year risk of CHD reduces predicted risk and improves patients' modifiable risk factor profile. Int J Clin Pract 62: 1484–1498.
73. Bonds DE, Hogan PE, Bertoni AG, Chen H, Clinch CR, et al. (2009) A multifaceted intervention to improve blood pressure control: The Guideline Adherence for Heart Health (GLAD) study. Am Heart J 157: 278–284.
74. Fiscella K, Volpe E, Winters P, Brown M, Idris A, et al. (2010) A novel approach to quality improvement in a safety-net practice: concurrent peer review visits. Journal of the National Medical Association 102: 1231–1236.
75. Kogan JR, Reynolds EE, Shea JA (2003) Effectiveness of report cards based on chart audits of residents' adherence to practice guidelines on practice performance: a randomized controlled trial. Teach Learn Med 15: 25–30.
76. Luders S, Schrader J, Schmieder RE, Smolka W, Wegscheider K, et al. (2010) Improvement of hypertension management by structured physician education and feedback system: cluster randomized trial. Eur J Cardiovasc Prev Rehabil 17: 271–279.
77. Mitchell E, Sullivan F, Watt G, Grimshaw JM, Donnan PT (2004) Using electronic patient records to inform strategic decision making in primary care. Stud Health Technol Inform 107: 2–61.
78. Sondergaard J, Hansen DG, Aarslev P, Munck AP (2006) A multifaceted intervention according to the Audit Project Odense method improved secondary prevention of ischemic heart disease: A randomised controlled trial. Fam Pract 23: 198–202.
79. Verstappen WHJM, van der Weijden T, Sijbrandij J, Smeele I, Hermsen J, et al. (2003) Effect of a Practice-Based Strategy on Test Ordering Performance of Primary Care Physicians: A Randomized Trial. JAMA 289: 2407–2412.
80. Amado GE, Pujol RE, Pacheco HV, Borras JM, Group A (2011) Knowledge and adherence to antihypertensive therapy in primary care: results of a randomized trial. Gaceta Sanitaria 25: 62–67.
81. Anguita Sánchez M, Jiménez-Navarro M, Crespo M, Alonso-Pulpón L, de Teresa E, et al. (2010) Using the Teamlet Model to Improve Chronic Care in an Academic Primary Care Practice. Revista española de cardiologia 63: 677–685.
82. De Muylder R, Tonglet R, Nackers F, Boland B (2005) Randomised evaluation of a specific training of general practitioners in cardiovascular prevention. Acta Cardiol 60: 199–205.
83. Feder G, Griffiths C, Eldridge S, Spence M (1999) Effect of postal prompts to patients and general practitioners on the quality of primary care after a coronary event (POST): Randomised controlled trial. BMJ 318: 1522–1526.
84. Feldman RD, Zou GY, Vandervoort MK, Wong CJ, Nelson SA, et al. (2009) A simplified approach to the treatment of uncomplicated hypertension: a cluster randomized, controlled trial. Hypertension 53: 646–653.
85. Goldberg HI, Wagner EH, Fihn SD, Martin DP, Horowitz CR, et al. (1998) A randomized controlled trial of CQI teams and academic detailing: can they alter compliance with guidelines? Jt Comm J Qual Improv 24: 130–142.
86. Jafar TH, Hatcher J, Poulter N, Islam M, Hashmi S, et al. (2009) Community-based interventions to promote blood pressure control in a developing country: a cluster randomized trial. Ann Intern Med 151: 593–601.
87. Kasje WN, Denig P, Stewart RE, de Graeff PA, Haaijer-Ruskamp FM (2006) An educational programme for peer review groups to improve treatment of chronic heart failure and diabetes mellitus type 2 in general practice. J Eval Clin Pract 12: 613–621.
88. Kiessling A, Henriksson P (2002) Efficacy of case method learning in general practice for secondary prevention in patients with coronary artery disease: Randomised controlled study. BMJ 325: 877–880.
89. Langham J, Tucker H, Sloan D, Pettifer J, Thom S, et al. (2002) Secondary prevention of cardiovascular disease: A randomised trial of training in information management, evidence-based medicine, both or neither: The PIER trial. Br J Gen Pract 52: 818–824.
90. Levine DA, Funkhouser EM, Houston TK, Gerald JK, Johnson-Roe N, et al. (2011) Improving care after myocardial infarction using a 2-year internet-delivered intervention: the Department of Veterans Affairs myocardial infarction-plus cluster-randomized trial. Arch Intern Med 171: 1910–1917.
91. Martensson J, Stromberg A, Dahlstrom U, Karlsson JE, Fridlund B (2005) Patients with heart failure in primary health care: effects of a nurse-led intervention on health-related quality of life and depression. Eur J Heart Fail 7: 393–403.
92. McAlister FA, Fradette M, Majumdar SR, Williams R, Graham M, et al. (2009) The enhancing secondary prevention in coronary artery disease trial. CMAJ 181: 897–904.
93. McBride P, Underbakke G, Plane MB, Massoth K, Brown RL, et al. (2000) Improving prevention systems in primary care practices: the Health Education and Research Trial (HEART). J Fam Pract 49: 115–125.
94. Mehler PS, Krantz MJ, Lundgren RA, Estacio RO, MacKenzie TD, et al. (2005) Bridging the quality gap in diabetic hyperlipidemia: A practice-based intervention. Am J Med 118: 1414.e1413–1414.e1419.
95. Simon SR, Majumdar SR, Prosser LA, Salem-Schatz S, Warner C, et al. (2005) Group versus individual academic detailing to improve the use of antihypertensive medications in primary care: a cluster-randomized controlled trial. Am J Med 118: 521–528.
96. Zillich AJ, Ackermann RT, Stump TE, Ambuehl RJ, Downs SM, et al. (2008) An evaluation of educational outreach to improve evidence-based prescribing in Medicaid: A cautionary tale. J Eval Clin Pract 14: 854–860.
97. Holbrook A, Pullenayegum E, Thabane L, Troyan S, Foster G, et al. (2011) Shared electronic vascular risk decision support in primary care: Computerization of Medical Practices for the Enhancement of Therapeutic Effectiveness (COMPETE III) randomized trial. Arch Intern Med 171: 1736–1744.
98. Lapointe F, Lepage S, Larriv'e L, Maheux P (2006) Surveillance and treatment of dyslipidemia in the post-infarct patient: can a nurse-led management approach make a difference? Can J Cardiol 22: 761–767.
99. Montgomery AA, Fahey T, Peters TJ (2003) A factorial randomised controlled trial of decision analysis and an information video plus leaflet for newly diagnosed hypertensive patients. Br J Gen Pract 53: 446–453.
100. Muhlhauser I, Sawicki PT, Didjurgeit U, Jurgens V, Trampisch HJ, et al. (1993) Evaluation of a structured treatment and teaching programme on hypertension in general practice. Clin Exp Hypertens 15: 125–142.
101. Munoz MA, Vila J, Cabanero M, Rebato C, Subirana I, et al. (2007) Efficacy of an intensive prevention program in coronary patients in primary care, a randomised clinical trial. Int J Cardiol 118: 312–320.
102. Oddone EZ, Weinberger M, Giobbie-Hurder A, Landsman P, Henderson W (1999) Enhanced access to primary care for patients with congestive heart failure. Veterans Affairs Cooperative Study Group on Primary Care and Hospital Readmission. Eff Clin Pract 2: 201–209.
103. O'Hare JP, Raymond NT, Mughal S, Dodd L, Hanif W, et al. (2004) Evaluation of delivery of enhanced diabetes care to patients of South Asian ethnicity: The United Kingdom Asian Diabetes Study (UKADS). Diabet Med 21: 1357–1365.
104. Saunders LD, Irwig LM, Gear JS, Ramushu DL (1991) A randomized controlled trial of compliance improving strategies in Soweto hypertensives. Med Care 29: 669–678.
105. Wood DA, Kotseva K, Connolly S, Jennings C, Mead A, et al. (2008) Nurse-coordinated multidisciplinary, family-based cardiovascular disease prevention programme (EUROACTION) for patients with coronary heart disease and asymptomatic individuals at high risk of cardiovascular disease: a paired, cluster-randomised controlled trial. Lancet 371: 1999–2012.
106. Grover SA, Lowensteyn I, Joseph L, Kaouache M, Marchand S, et al. (2007) Patient knowledge of coronary risk profile improves the effectiveness of dyslipidemia therapy: the CHECK-UP study: a randomized controlled trial. Arch Intern Med 167: 2296–2303.
107. McManus RJ, Mant J, Roalfe A, Oakes RA, Bryan S, et al. (2005) Targets and self monitoring in hypertension: randomised controlled trial and cost effectiveness analysis. BMJ 331: 493.
108. McManus RJ, Mant J, Bray EP, Holder R, Jones MI, et al. (2010) Telemonitoring and self-management in the control of hypertension (TASMINH2): a randomised controlled trial. Lancet 376: 163–172.
109. Parati G, Omboni S, Albini F, Piantoni L, Giuliano A, et al. (2009) Home blood pressure telemonitoring improves hypertension control in general practice. The TeleBPCare study. J Hypertens 27: 198–203.
110. Baig AA, Mangione CM, Sorrell-Thompson AL, Miranda JM (2010) A randomized community-based intervention trial comparing faith community nurse referrals to telephone-assisted physician appointments for health fair participants with elevated blood pressure. Journal of General Internal Medicine 25: 701–709.
111. Bogden PE, Koontz LM, Williamson P, Abbott RD (1997) The physician and pharmacist team: An effective approach to cholesterol reduction. J Gen Intern Med 12: 158–164.
112. Community Pharmacy Medicines Management Projekt Evaluation Team (2007) The MEDMAN study: A randomized controlled trial of community pharmacy-led medicines management for patients with coronary heart disease. Fam Pract 24: 189–200.
113. Bond CM, Fish A, Porteous TH, Reid JP, Scott A, et al. (2007) A randomised controlled trial of the effects of note-based medication review by community pharmacists on prescribing of cardiovascular drugs in general practice. Int J Pharm Pract 15: 39–46.
114. Borenstein JE, Graber G, Saltiel E, Wallace J, Ryu S, et al. (2003) Physician-pharmacist comanagement of hypertension: a randomized, comparative trial. Pharmacotherapy 23: 209–216.
115. Carter BL, Ardery G, Dawson JD, James PA, Bergus GR, et al. (2009) Physician and pharmacist collaboration to improve blood pressure control. Arch Intern Med 169: 1996–2002.
116. Fihn SD, Bucher JB, McDonell M, Diehr P, Rumsfeld JS, et al. (2011) Collaborative care intervention for stable ischemic heart disease. Arch Intern Med 171: 1471–1479.
117. Hunt JS, Siemienczuk J, Pape G, Rozenfeld Y, MacKay J, et al. (2008) A randomized controlled trial of team-based care: impact of physician-pharmacist collaboration on uncontrolled hypertension. J Gen Intern Med 23: 1966–1972.
118. Jolly K, Bradley F, Sharp S, Smith H, Thompson S, et al. (1999) Randomised controlled trial of follow up care in general practice of patients with myocardial infarction and angina. Final results of the Southampton heart integrated care project (SHIP). BMJ 318: 706–711.
119. Kaczorowski J, Chambers LW, Dolovich L, Paterson JM, Karwalajtys T, et al. (2011) Improving cardiovascular health at population level: 39 community cluster randomised trial of Cardiovascular Health Awareness Program (CHAP). BMJ d442.
120. Katon WJ, Lin EHB, Korff MV, Ciechanowski P, Ludman EJ, et al. (2010) Collaborative care for patients with depression and chronic illnesses. N Engl J Med 363: 2611–2620.

121. Lobo CM, Frijling BD, Hulscher ME, Bernsen RM, Braspenning JC, et al. (2002) Improving quality of organizing cardiovascular preventive care in general practice by outreach visitors: a randomized controlled trial. Prev Med 35: 422–429.
122. McInnes GT, McGhee SM (1995) Delivery of care for hypertension. J Hum Hypertens 9: 429–433.
123. Simpson SH, Majumdar SR, Tsuyuki RT, Lewanczuk RZ, Spooner R, et al. (2011) Effect of adding pharmacists to primary care teams on blood pressure control in patients with type 2 diabetes: A randomized controlled trial. Diabetes Care 34: 20–26.

Tabellarischer Lebenslauf

PERSÖNLICHE ANGABEN

Dr. rer. nat. Susanne Unverzagt

Geboren am 15.02.1969 in Wernigerode

3 Kinder (24, 19 und 10 Jahre)

BERUFSERFAHRUNG

02 / 2007 bis heute — Wissenschaftliche Mitarbeiterin

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

- Projektleitung und statistische Auswertung für systematische Übersichtsarbeiten
- Methodische Leitung der Evidenzaufarbeitung für S3-Leitlinien
- Biometrische Beratungen
- Betreuung von medizinischen Doktorarbeiten
- Seminare und Vorlesungen zu Biometrie und Epidemiologie
- Gutachterin für medizinische Fachzeitschriften

10 / 2002 bis 03 / 2005 — Biometrikerin

Koordinierungszentrum für Klinische Studien Halle (Martin-Luther Universität Halle/Wittenberg)

- Projektmanagement für Interventions- und Diagnosestudien
- Abfassung von Arbeitsabläufen (SOPs)

01 / 2001 bis 04 / 2002 — Wissenschaftliche Mitarbeiterin

Institut für Mathematik und Informatik und Klinik für Neurologie (Ernst-Moritz-Arndt-Universität Greifswald)

11 / 1994 – 12 / 2000 — Wissenschaftliche Mitarbeiterin

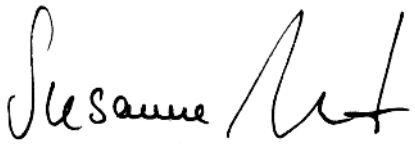
Geographisches Institut (Ernst-Moritz-Arndt-Universität Greifswald)

AUSBILDUNG UND SPEZIELLE QUALIFIKATIONEN

Seit 2012 — Kurse zu Hochschuldidaktik und wissenschaftlichen Arbeitsmethoden (Martin-Luther Universität Halle/Wittenberg)

Seit 2008 — Kurse zu systematischen Übersichtsarbeiten, Literaturrecherche und Evidenzbewertung für Interventions- und Diagnosestudien (Deutsches Cochrane Zentrum Freiburg)

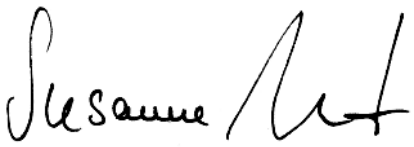
2002-2012	**Postgraduales und Masterstudium** Ruprechts-Karls-Universität Heidelberg ▪ Master of Science (Medical Biometry/Biostatistics) (Gesamtnote 1,8)
1994 – 2000	**Promotion** Ernst-Moritz-Arndt-Universität Greifswald ▪ Dr. rer. nat. (Gesamtnote summa cum laude)
1987 – 1994	**Mathematikstudium** Ernst-Moritz-Arndt-Universität Greifswald ▪ Diplommathematiker (Prädikat gut)
1983 – 1987	**Gymnasium** Erweiterte Oberschule „Gerhard-Hauptmann“ Wernigerode und „Bertold-Brecht“ Halberstadt ▪ Abitur
1975 – 1983	**Grund-und Gesamtschule** August-Hermann-Francke, Wernigerode



Halle, den 04.05.2015 Dr. Susanne Unverzagt

Selbstständigkeitserklärung

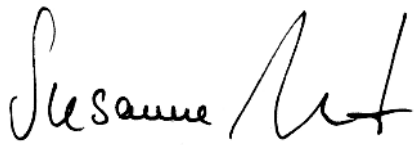
Ich erkläre hiermit, die Arbeit selbständig geschrieben und keine anderen als die angegebenen Quellen benutzt zu haben.



Halle, den 04.05.2015 Dr. Susanne Unverzagt

Erklärung über frühere Habilitationsversuche

Diese Arbeit habe ich im Rahmen meines ersten Habilitationsversuches geschrieben. Ich habe diese Arbeit ausschließlich an der Medizinischen Fakultät der Martin-Luther-Universität Halle-Wittenberg als Habilitationsschrift eingereicht.



Halle, den 04.05.2015 Dr. Susanne Unverzagt