

# TECHNISCHE UNIVERSITÄT MÜNCHEN

Institut für Allgemeinmedizin

(Direktor: Univ.-Prof. Dr. Antonius Schneider)

## **Inhalte, Qualität und Ergebnisse von randomisierten kontrollierten Studien der universitären Allgemeinmedizin in Deutschland – ein systematischer Review von 2000 bis 2014**

Stefan Philipp Heinmüller

Vollständiger Abdruck der von der Fakultät für Medizin  
der Technischen Universität München zur Erlangung des akademischen Grades eines  
Doktors der Medizin genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Jürgen Schlegel

Prüfer der Dissertation:

1. apl. Prof. Dr. Klaus Linde
2. Univ.-Prof. Dr. Antonius Schneider

Die Dissertation wurde am 08.02.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 21.02.2018 angenommen.



*Meinen Großeltern  
Erika und Dr. Werner Heinmüller  
aus Mansbach*



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung und Zielsetzung</b> .....	<b>1</b>
1.1	Bedeutung der Hausarztmedizin .....	1
1.2	Notwendigkeit der Forschung in der Allgemeinmedizin sowie deren universitärer Anbindung .....	2
1.3	Historische Entwicklung der akademischen Allgemeinmedizin in Deutschland.....	3
1.4	Herausforderungen der universitären Allgemeinmedizin .....	5
<b>2</b>	<b>Methodik</b> .....	<b>9</b>
2.1	Protokoll und Regelwerk.....	9
2.2	Einschlusskriterien .....	9
2.2.1	Publikationszeitraum.....	9
2.2.2	Studiendesign .....	9
2.2.3	Affiliation zur deutschen universitären Allgemeinmedizin .....	10
2.3	Literatursuche .....	10
2.3.1	2000 bis 2010: Publikationen aus Vorgängerdissertation.....	11
2.3.2	2010 bis 2014: Datenbanksuche Scopus und Durchsicht von Publikationslisten .....	12
2.3.3	Durchsicht von Literaturangaben (engl. reference tracking) .....	14
2.4	Selektionsprozess.....	15
2.4.1	2000 bis 2010: Selektion der Publikationen aus Vorgängerdissertation .....	15
2.4.2	2010 bis 2014: Selektion der Publikationen aus Datenbanksuche Scopus und Durchsicht von Publikationslisten.....	16
2.5	Extraktion von Studieninformationen .....	17
2.5.1	Extraktionsprozess – praktisches Vorgehen .....	17
2.5.2	Formale Merkmale und Finanzierung.....	20
2.5.3	Thematische Einordnung .....	21
2.5.4	Studienbeteiligte und Studienteilnehmer .....	21
2.5.5	Details zum Studiendesign .....	23
2.5.6	Ergebnisse der untersuchten Studien.....	24

<b>2.6</b>	<b>Beurteilung von Verzerrungsrisiken</b> .....	<b>24</b>
2.6.1	„Random sequence generation“ .....	25
2.6.2	„Allocation concealment“ .....	26
2.6.3	„Blinding of participants and personnel“ .....	28
2.6.4	„Blinding of outcome assessment“ .....	30
2.6.5	„Incomplete outcome data“ .....	31
2.6.6	„Selective reporting“ .....	33
<b>2.7</b>	<b>Zusammenfassung und Beurteilung von Studienergebnissen</b> .....	<b>33</b>
2.7.1	„Vote count“ .....	33
2.7.2	Metaanalyse .....	35
<b>3</b>	<b>Ergebnisse</b> .....	<b>37</b>
<b>3.1</b>	<b>Literatursuche und Selektionsprozess</b> .....	<b>37</b>
3.1.1	Publikationen aus Vorgängerdissertation, Scopussuche, „reference tracking“ .....	37
3.1.2	Publikationslisten .....	37
3.1.3	Selektionsprozess .....	38
3.1.4	Nachträglicher Studienausschluss .....	39
<b>3.2</b>	<b>Merkmale der eingeschlossenen Studien</b> .....	<b>42</b>
3.2.1	Zeitpunkt der Veröffentlichung, Publikationssprache und Autorenschaft .....	42
3.2.2	Beteiligte Abteilungen .....	45
3.2.3	Finanzierung .....	46
3.2.4	Untersuchte Fragestellungen .....	48
3.2.5	Informationen zum Studiendesign .....	57
3.2.6	Orte der Studiendurchführung .....	63
3.2.7	Beteiligte und Teilnehmer der untersuchten Studien .....	64
<b>3.3</b>	<b>Verzerrungsrisiken der untersuchten Studien</b> .....	<b>69</b>
3.3.1	Verzerrungsrisiken der clusterrandomisierten Studien .....	69
3.3.2	Verzerrungsrisiken der „klassischen“ randomisierten Studien .....	70
<b>3.4</b>	<b>Zusammenfassung und Beurteilung der Studienergebnisse</b> .....	<b>73</b>
3.4.1	„Vote count“ .....	73
3.4.2	Metaanalyse .....	75

<b>4</b>	<b>Diskussion .....</b>	<b>77</b>
4.1	<b>Zusammenfassung der Hauptergebnisse .....</b>	<b>77</b>
4.2	<b>Methodenkritik .....</b>	<b>78</b>
4.2.1	Literatursuche .....	78
4.2.2	Erste Arbeit dieser Art – Stärke oder Schwäche?.....	78
4.2.3	Einschlusskriterien .....	79
4.2.4	Hürden und Kritikpunkte der Studienanalyse .....	80
4.3	<b>Interpretation.....</b>	<b>82</b>
4.3.1	RCTs der deutschen universitären Allgemeinmedizin – Entwicklung und Vergleich .....	82
4.3.2	Clusterrandomisierte Studien der universitären Allgemeinmedizinforschung	84
4.3.3	Inhaltliche Einordnung der analysierten Studien.....	86
4.3.4	Methodik der analysierten Studien .....	87
4.3.5	Effekte der untersuchten Interventionen.....	89
4.4	<b>Schlussfolgerung .....</b>	<b>89</b>
<b>5</b>	<b>Zusammenfassung.....</b>	<b>91</b>
	<b>Abbildungsverzeichnis.....</b>	<b>i</b>
	<b>Tabellenverzeichnis.....</b>	<b>iii</b>
	<b>Abkürzungsverzeichnis.....</b>	<b>v</b>
	<b>Literaturverzeichnis .....</b>	<b>xi</b>
	Verzeichnis aller Text- und Studienreferenzen .....	xi
	Übersicht aller Studienreferenzen .....	xxviii
	<b>Anhang .....</b>	<b>xxix</b>
	<b>Publikationsverzeichnis .....</b>	<b>xxxviii</b>
	<b>Danksagung .....</b>	<b>xxxix</b>



# 1 Einleitung und Zielsetzung

## 1.1 Bedeutung der Hausarztmedizin

Die Ergebnisse von Versorgungsforschungsstudien sprechen dafür, dass Gesundheitssysteme, die auf einer leistungsfähigen und flächendeckend verfügbaren Hausarztmedizin fußen, viele Vorteile mit sich bringen. Hierzu zählen ein besseres Gesundheitsniveau innerhalb der Bevölkerung, eine höhere Patientenzufriedenheit, geringere Gesundheitsausgaben sowie die Verhinderung von Über- und Fehlversorgung [Starfield, 1994]. Darüber hinaus gibt es Hinweise auf einen positiven Zusammenhang zwischen der Verfügbarkeit einer primärärztlichen Versorgung und der Gesamtmortalität sowie der Lebenserwartung der Menschen [Shi, 1994]. Viele Länder haben die genannten Vorteile einer breiten Primärversorgung erkannt und deshalb eine Stärkung derselben – inklusive entsprechender universitärer Einrichtungen – angestrebt. Adäquat umsetzen ließen sich solche Vorhaben bisher allerdings hauptsächlich in Ländern, die aufgrund staatlich stark beeinflusster Gesundheitssysteme über die nötigen Steuerungsmechanismen verfügen, um solche Veränderungen in der Versorgungsstruktur vornehmen zu können; hierfür sind die Niederlande, UK, Skandinavien und Kanada gute Beispiele [Donner-Banzhoff, 2016].

In Deutschland ist die flächendeckende allgemeinmedizinische Versorgung historisch gewachsen und nimmt eine bedeutende gesellschaftliche Funktion wahr: Für einen großen Teil der Bevölkerung ist der „Hausarzt<sup>1</sup> am Ort“ nach wie vor die erste Anlaufstelle bei medizinischen Problemen. So suchen derzeit knapp 60 Prozent der Menschen, die ärztlichen Rat in Anspruch nehmen möchten, zuerst einen Allgemeinarzt auf [Mader, 2014, S. 24]. Dieser behandelt rund 90 Prozent sämtlicher Beratungsprobleme eigenverantwortlich (ohne Überweisung an einen Spezialisten) [Braun et al., 1986 in Schmittziel, 2002, S. 74]. Somit wird ein beträchtlicher Anteil der Deutschen, die aufgrund gesundheitlicher Beschwerden einen Arzt kontaktieren, bis zum Abklingen derselben ausschließlich durch einen Hausarzt betreut.

Fachärzte für Allgemeinmedizin stellen in Deutschland nach den Internisten die zweitgrößte Arztgruppe mit Fachgebietsbezeichnung unter allen berufstätigen Ärzten dar [BÄK, 2015, Abb. 3]. Würde man jedoch die „Hausarztmedizin“ als eigene Fachrichtung betrachten,

---

<sup>1</sup> Aus Gründen der besseren Lesbarkeit wird in der gesamten vorliegenden Arbeit auf die gleichzeitige Verwendung männlicher und weiblicher Sprachformen verzichtet. Sämtliche Personenbezeichnungen gelten gleichermaßen für beiderlei Geschlecht.

wobei dann neben den Allgemeinmedizinern auch hausärztlich tätige Internisten sowie praktische Ärzte berücksichtigt werden müssten, so würden diese „Hausarztmediziner“ – abgesehen von den Ärzten ohne Gebietsbezeichnung – die größte Arztgruppe innerhalb aller berufstätigen Ärzte in Deutschland darstellen.

### **1.2 Notwendigkeit der Forschung in der Allgemeinmedizin sowie deren universitärer Anbindung**

Die moderne Schulmedizin fußt auf wissenschaftlichen Erkenntnissen, die durch systematische Forschung gewonnen werden müssen. Dies gilt für alle medizinischen Fachdisziplinen – einschließlich der Allgemeinmedizin – gleichermaßen. Entsprechende, qualitativ hochwertige Forschungsprojekte werden in der Regel durch Universitäten organisiert und ausgeführt [BMBF, 2016], weil diese – aufgrund deren historischer Entwicklung als Zentren der Forschung – über eine ausreichende Infrastruktur, die nötige fachliche Expertise und staatliche Fördergelder verfügen.

Kritische Stimmen behaupten bis heute, dass in der Allgemeinmedizin keine bzw. weniger Forschung nötig sei, weil sich die Allgemeinmedizin zum einen Ergebnissen anderer Fächer (wie z. B. der Inneren Medizin, Chirurgie oder Pädiatrie) bedienen könne, zum anderen, weil der Schweregrad der in der Hausarztmedizin vorkommenden Erkrankungen weit unter dem der in Universitätskliniken behandelten Krankheiten liege, sodass der Erforschung allgemeinmedizinischer Fragestellungen grundsätzlich nur eine untergeordnete Priorität zukommen könne. Dies sind jedoch keine hinreichenden Rechtfertigungen für eine geringere Forschungsaktivität in der Allgemeinmedizin (im Vergleich zu anderen medizinischen Fächern), da sich die Notwendigkeit allgemeinmedizinischer Forschung allein schon aus der mit Abstand höchsten Anzahl an Patientenkontakten (im Vergleich zu anderen Disziplinen) ergibt. Green et al. [2001] schätzten, dass von 1000 in den USA lebenden Menschen durchschnittlich 113 jeden Monat einen Hausarzt aufsuchen, während nur eine Person in einem Universitätsklinikum behandelt wird. Dieser Schätzung zufolge übersteigt die Anzahl der allgemeinmedizinischen Patientenkontakte diejenigen der Universitätsmedizin um mehr als den Faktor 100.

Weiterhin zeigt die von Green et al. [2001] durchgeführte Schätzung, wie stark das Patientengut der Universitätskliniken vorselektiert ist und dass es das hausärztliche Patientenspektrum deshalb nur schlecht repräsentieren kann. Folglich ist die Übertragung von in (Universitäts-)Kliniken gewonnenen Erkenntnissen auf den hausärztlichen Bereich problematisch [DEGAM-Zukunftspositionen, 2012]. Nichts desto trotz findet der überwiegende Teil klinischer Studien nach wie vor an Universitätskliniken statt; somit wird

den Besonderheiten der Allgemeinmedizin (wie beispielsweise dem unausgelesenen Patientenkollektiv und der Arbeit im Niedrigprävalenzbereich) nicht in ausreichendem Maß Rechnung getragen.

Aus den genannten Gründen sollten vermehrt qualitativ hochwertige klinisch-hausärztliche Forschungsprojekte angestrebt werden, deren Realisierung vermutlich nur durch eine starke akademische Allgemeinmedizin bewerkstelligt werden kann.

### **1.3 Historische Entwicklung der akademischen Allgemeinmedizin in Deutschland**

Wie in 1.1 erläutert, hat die Allgemeinmedizin in Deutschland eine bedeutende Stellung für das Gesundheitssystem, die Bevölkerung sowie die Ärzteschaft inne. Daraus resultiert auch die Notwendigkeit eigenständiger allgemeinmedizinischer Forschung (1.2). Vor diesem Hintergrund verwundert es, dass die Allgemeinmedizin auf universitärer Ebene lange Zeit praktisch nicht existierte, während die anderen klassischen medizinischen Disziplinen spätestens seit Beginn des 20. Jahrhunderts (häufig aber schon viel länger) ein fester Bestandteil fast aller medizinischer Fakultäten waren.

Erst 1976 wurde in Hannover mit der Vergabe des ersten allgemeinmedizinischen Lehrstuhls [Hummers-Pradier, 2006, Tab. 1] der Grundstein der akademischen Allgemeinmedizin gelegt.<sup>1</sup> Die Institutionalisierung derselben machte jedoch auch danach nur bescheidene Fortschritte. Als die Kultusministerkonferenz 1989 beschloss, bis Anfang der 1990er Jahre an allen deutschen medizinischen Hochschulen ordentliche Lehrstühle für Allgemeinmedizin einzurichten [Josupeit, 2002, S. 5], waren seit der Gründung des ersten allgemeinmedizinischen Lehrstuhls bereits 13 Jahre vergangen. Zudem blieb der genannte Beschluss praktisch ohne Folgen: Im Jahr 1999 existierten immer noch an nur 6 der 37 medizinischen Fakultäten in Deutschland Lehrstühle für Allgemeinmedizin [Kochen, 1999].

Erst zu Beginn des 21. Jahrhunderts nahm die Akademisierung der Allgemeinmedizin an Fahrt auf – hierzu leistete vermutlich die Aufmerksamkeit, die das Fach durch den gefürchteten, in Medien und Politik viel diskutierten „Hausärztemangel“ [KBV, 2016] erhalten hat, einen nicht unerheblichen Beitrag. So gab es im Jahr 2006 an 13 [Hummers-

---

<sup>1</sup> Der erste allgemeinmedizinische Lehrauftrag (ohne eigenständigen Lehrstuhl) war bereits 1966 durch die Universität Freiburg vergeben worden [Hummers-Pradier, 2006, Tab. 1].

Pradier, 2006], im Jahr 2014 immerhin an 20 der 37 medizinischen Fakultäten eigenständige Lehrstühle für Allgemeinmedizin [Schneider, 2016].<sup>1</sup>

Die Gründe für die beschriebene zögerliche Entwicklung der forschenden Allgemeinmedizin lassen sich nicht genau nachvollziehen. Fakt ist aber, dass es gegen deren Anbindung an Universitäten zahlreiche Vorbehalte gab [Gerlach, 2016]. Als die Möglichkeit der allgemeinmedizinischen Habilitation geschaffen werden sollte, stieß dies teilweise sogar in den eigenen Reihen auf Ablehnung [Fischer, 2016]. Häufiger kamen Bedenken und Widerstände jedoch aus etablierten Fachdisziplinen innerhalb der Universitätsmedizin, die einer Präsenz der Allgemeinmedizin in Forschung und Lehre vielerorts kritisch gegenüberstanden [Mader, 2016]. Zeitweise wurde sogar die Existenzberechtigung der gesamten Hausarztmedizin in Frage gestellt: Manche Kritiker vertraten die Meinung, dass der praktische Arzt nicht mit dem medizinischen Fortschritt mithalten könne und deshalb seinem Aufgabengebiet nicht mehr gewachsen sei. Die Patientenversorgung, so hieß es, solle gänzlich an Kliniken und Fachpraxen abgegeben werden [Häussler, 2016]. Darüber hinaus gibt es – wie bereits in 1.2 angesprochen – bis heute Stimmen, die allgemeinmedizinische Forschung mit der Begründung ablehnen, dass sich diese der Forschungsergebnisse anderer Disziplinen (wie z. B. der Inneren Medizin, Chirurgie oder Pädiatrie) bedienen könne.

Zusammenfassend kann gesagt werden, dass die allgemeinmedizinische Forschung in Deutschland lange vernachlässigt wurde. Noch 1983 behauptete Schwartz [1983]: „Der Rückstand der Forschung in der Allgemeinmedizin ist ein Defizit für die Medizin überhaupt. Er ist nicht das Ergebnis objektiver Unmöglichkeit von Forschung in der Allgemeinmedizin, sondern Ergebnis unterlassener Forschungsinvestitionen. Forschung in der Allgemeinmedizin ist ein gemeinsames Anliegen der medizinisch-wissenschaftlichen Gemeinschaft“. Zwar hat sich die Situation der akademischen Allgemeinmedizin in Deutschland seit den 1980er Jahren verbessert (z. B. durch die Neugründung von Lehrstühlen); in Bezug auf Mitarbeiterzahlen, Finanzmittel, Forschungsoutput etc. kann sie dennoch bis heute weder mit den meisten anderen medizinischen Fachdisziplinen in Deutschland noch mit der forschenden Allgemeinmedizin vieler anderer Länder wie Australien, Kanada, den Niederlanden, UK oder USA mithalten [Glanville et al., 2011].

---

<sup>1</sup> Die o. g. Zahlen beziehen sich ausschließlich auf vollwertige Lehrstühle für Allgemeinmedizin, die mit ordentlichen Universitätsprofessoren besetzt sind. Nicht berücksichtigt werden hingegen allgemeinmedizinische Lehrbereiche ohne Lehrstuhl (diese werden meist ehrenamtlich von engagierten niedergelassenen Hausärzten organisiert – wie z. B. an der Universität Regensburg [Richter-Kuhlmann et al., 2015]).

### 1.4 Herausforderungen der universitären Allgemeinmedizin

Doch warum hinkt die allgemeinmedizinische Forschung in Deutschland – wie in 1.3 dargestellt – so weit hinterher? Die Gründe hierfür sind mannigfaltig. Hummers-Pradier [2006] nennt die historische Entwicklung des Fachs mit seiner primären Ausrichtung auf die praktische ärztliche Tätigkeit als einen möglichen Grund, weshalb sich nur schrittweise ein wissenschaftliches Selbstverständnis entwickelt hat.

Darüber hinaus hat die Allgemeinmedizin in Deutschland eine weitere Sonderstellung inne: Während die meisten medizinischen Fächer entweder ausschließlich im stationären Bereich (z. B. Herzchirurgie, Intensivmedizin) oder in der stationären und ambulanten Versorgung gleichermaßen (z. B. Pädiatrie, Gynäkologie, Dermatologie, Hals-Nasen-Ohren Heilkunde etc.) beheimatet sind, ist die Allgemeinmedizin einzig im ambulanten und *nicht* im stationären Setting angesiedelt. In der zum überwiegenden Teil in (Universitäts-)Kliniken stattfindenden klinischen Forschung sind allgemeinmedizinische Fragestellungen somit unterrepräsentiert.

Eigenständige hausärztliche Forschungsprojekte zu klinischen Fragestellungen können, da die Allgemeinmedizin – wie erwähnt – nicht in Kliniken vertreten ist, praktisch nur in Hausarztpraxen durchgeführt werden. Viele Hausärzte in Deutschland haben allerdings kaum freie Kapazitäten für die Forschung, da sie als Selbstständige in Einzel- oder kleinen Gemeinschaftspraxen unter großem bürokratischen Aufwand, ökonomischem Druck und einer hohen zeitlichen Arbeitslast (bei der Anzahl der pro Tag behandelten Patienten hält Deutschland sogar den „Europarekord“ [Hummers-Pradier, 2006]) leiden. Weiterhin wird teilweise unter Hausärzten die Meinung vertreten, Forschungsprojekte in der Praxis würden bei den Patienten nur auf geringe Akzeptanz stoßen und zu einer Abwanderung derselben führen, was in einem marktorientierten ambulanten Gesundheitssystem finanzielle Einbußen für den Arzt nach sich ziehen würde [Hummers-Pradier et al., 2012].

Daneben gibt es noch zahlreiche weitere Herausforderungen, denen die akademische Allgemeinmedizin in Deutschland gegenübersteht. Beispielsweise dienen allgemeinmedizinische Studien häufig der Überprüfung, ob Forschungsergebnisse, die bereits durch Studien an Universitätskliniken gewonnen wurden, auch auf den ambulanten Bereich übertragbar sind bzw. wie diese in der Primärversorgung implementiert werden können. Doch obwohl solche und ähnliche Fragestellungen höchst praxisrelevant sind, werden sie von Kritikern häufig als „nicht innovativ genug“ oder als „nicht in ausreichendem Maße einem naturwissenschaftlichen Erkenntnisgewinn dienend“ angesehen.

Angesichts der o. g. Forschungsschwerpunkte sind Fragestellungen zu neuen, teuren Medikamenten oder invasiven Interventionen in der Allgemeinmedizin eher die Ausnahme. Dies leitet zu einem weiteren Thema über, das bei allgemeinmedizinischen Studien häufig eine Hürde darstellt: die Finanzierung. Pharma- und Medizintechnikindustrie sind aus nachvollziehbaren Gründen nur wenig an der Erforschung preiswerter und wenig invasiver Therapien interessiert und scheiden somit als Förderer größtenteils aus. Folglich kommen für die Finanzierung allgemeinmedizinischer Studien in erster Linie öffentliche Gelder in Frage. Bei der Vergabe derselben, z. B. über die Deutsche Forschungsgemeinschaft (DFG), herrscht allerdings ein starker Konkurrenzdruck, wobei die in Deutschland besonders angesehene experimentelle Grundlagenforschung, die in einem Fach wie der Allgemeinmedizin praktisch nicht möglich ist, häufig die meisten Mittel zugesprochen bekommt [Hummers-Pradier, 2006, Hummers-Pradier et al., 2012]. Im Bereich der Versorgungs- und der klinischen Forschung, die für die Allgemeinmedizin besonders relevant sind, gelten hingegen randomisierte kontrollierte Studien (engl. randomised controlled trials = RCTs) als Königsweg [Kabisch et al., 2011]. Methodisch hochwertige RCTs zu klinisch relevanten Fragestellungen können für die Patientenversorgung von unmittelbarer Bedeutung sein [Gágyor et al., 2011]. Planung, Durchführung und Auswertung von randomisierten kontrollierten Studien sind jedoch besonders aufwendig und setzen eine ausreichende Infrastruktur sowie methodische und praktische Kenntnisse voraus. Das wiederum ist mit weitaus höheren Kosten verbunden als dies bei experimentellen Forschungsprojekten vergleichbarer Qualität<sup>1</sup> der Fall ist. Hierdurch wird die finanzielle Situation für allgemeinmedizinische Forschungsprojekte, die – wie oben erwähnt – bei der Vergabe von Finanzmitteln ohnehin nur unterproportional berücksichtigt werden, zusätzlich verschärft.

Doch auch abgesehen von diesen studienbezogenen Mitteln ist die finanzielle Ausstattung der akademischen Allgemeinmedizin in Deutschland, z. B. hinsichtlich allgemeinmedizinischer Universitätsabteilungen oder Forschungspraxis-Netzwerken, im Vergleich zu anderen Ländern wie Großbritannien, den USA oder den Niederlanden gering [Grätzel von Grätz, 2015, Hummers-Pradier et al., 2012, Hummers-Pradier et al., 2014].

Auch bei den allgemeinmedizinischen Instituten sind Defizite zu beklagen. So gibt es (Stand 2014) noch an etwa einem Drittel der medizinischen Fakultäten keinen Lehrstuhl für Allgemeinmedizin [Schneider, 2016]. Da sich die Zahl der allgemeinmedizinischen Lehrstühle in Deutschland jedoch in den letzten Jahren sehr positiv entwickelt hat und

---

<sup>1</sup> Beispielsweise könnten für diesen Vergleich die Impactfaktoren der Fachzeitschriften, in denen die jeweiligen Studienergebnisse publiziert werden, als Qualitätsindikatoren herangezogen werden.

davon auszugehen ist, dass in wenigen Jahren jede medizinische Fakultät über eine W3-Professur für Allgemeinmedizin verfügt [Donner-Banzhoff, 2016], wird an dieser Stelle auf eine weitere Ausführung verzichtet.

Zu erwähnen ist jedoch, dass die bestehenden universitären Institute für Allgemeinmedizin mit wenigen Ausnahmen klein sind [Hummers-Pradier, 2006]. Der Zustand der universitären Allgemeinmedizin aus dem Jahr 2006 – neuere Erhebungen hierzu sind leider nicht verfügbar – kann dazu dienen, die Extreme aufzuzeigen. Damals hatten zwölf der 36 medizinischen Fakultäten keinerlei wissenschaftliche Mitarbeiter<sup>1</sup> in der Allgemeinmedizin, weitere zwölf hatten maximal eine Vollzeitstelle und die maximale Anzahl an wissenschaftlichen Mitarbeitern lag bei fünf. An einigen Universitäten verfügten die allgemeinmedizinischen Abteilungen nicht einmal über eigene Räumlichkeiten [Baum et al., 2006]. Bei diesen Darstellungen muss natürlich berücksichtigt werden, dass sich die Rahmenbedingungen der akademischen Allgemeinmedizin in Deutschland seit 2006 verbessert haben.

Die meisten allgemeinmedizinischen Institute in Deutschland wurden erst in den letzten 15 Jahren gegründet und konnten aufgrund ihrer geringen Größe und mäßigen finanziellen Ausstattung nur wenige klinische Studien durchführen. Heute liegen deshalb nur begrenzt Erfahrungswerte in der Durchführung methodisch hochwertiger klinischer Studien vor. Zudem konnte kein akademischer Nachwuchs „heranwachsen“, weshalb es in Deutschland (Stand 2012) weniger als 30 habilitierte Allgemeinmediziner gibt [Hummers-Pradier et al., 2012]; im Jahr 2015 erfolgte nur eine einzige allgemeinmedizinische Habilitation [Habilitationen, 2015a, 2015b, 2015c, 2015d].<sup>2</sup>

Eine Forschungstätigkeit in einem allgemeinmedizinischen Universitätsinstitut scheint zudem nur für wenige Ärzte interessant zu sein, sodass sich die Suche nach geeignetem wissenschaftlichen Personal häufig schwierig gestaltet. Ein Grund hierfür ist, dass viele Universitäten ihre an den allgemeinmedizinischen Instituten angestellten Ärzte „nur“ als wissenschaftliche Mitarbeiter und nicht nach dem Tarifvertrag der Ärzte (TV-Ä) bezahlen, was für diese eine deutliche Gehaltseinbuße gegenüber einer Beschäftigung in Klinik oder Praxis darstellt [Hummers-Pradier et al., 2012]. Auch betrachten viele

---

<sup>1</sup> Wissenschaftliche Mitarbeiter sind hier als Angestellte im öffentlichen Dienst nach (dem damals gültigen) Bundes-Angestelltentarifvertrag (BAT) definiert. Andere Mitarbeiter, wie z. B. studentische Hilfskräfte, finden an dieser Stelle keine Berücksichtigung.

<sup>2</sup> Zum Vergleich: die insgesamt 675 medizinischen Habilitationen in Deutschland aus dem Jahr 2015 verteilten sich folgendermaßen auf die verschiedenen Fachdisziplinen: Chirurgie (142), Innere Medizin (116), Radiologie (49), Pädiatrie (37), Anästhesie (34), Neurologie (26), Augenheilkunde (25), Gynäkologie (24), Psychiatrie (17), Dermatologie (15), Urologie (15), HNO (9), Allgemeinmedizin (1), andere (165) [Habilitationen, 2015a, 2015b, 2015c, 2015d].

allgemeinmedizinische Weiterbildungsassistenten, die zukünftig eine Niederlassung anstreben, eine vorübergehende Anstellung in einem universitären Institut für Allgemeinmedizin als hinderlich, weil entsprechende Forschungszeiten (da nicht auf die Facharztweiterbildung anrechenbar) die Weiterbildungszeit in die Länge ziehen würden.

Folgender Aspekt scheint ebenfalls eine Rolle zu spielen: Während bei Klinikärzten eine wissenschaftliche Tätigkeit als Zukunftsinvestition betrachtet werden kann, die sich – zumindest auf lange Sicht – auch finanziell auszahlt (schließlich gilt die Habilitation an vielen großen Kliniken als Einstellungsvoraussetzung für gut bezahlte Ober- oder Chefarztpositionen), ist in der Allgemeinmedizin das Gegenteil der Fall. Hier verzögern lange Forschungszeiten den Zeitpunkt bis zur Niederlassung und haben – wenn überhaupt – nur einen geringen Einfluss auf die spätere Einkommenssituation eines Hausarztes. Finanziell lukrativer ist es für angehende Hausärzte, anstatt zu forschen, möglichst viele Zusatzbezeichnungen wie Akupunktur, Chirotherapie, Sportmedizin etc. zu erwerben, die ihnen später zusätzliche Abrechnungsmöglichkeiten eröffnen.

Zusammenfassend kann gesagt werden, dass die Bedingungen für hochwertige allgemeinmedizinische Forschung in Deutschland problematisch sind.

Die vorliegende Arbeit beabsichtigt, in Form eines systematischen Reviews, einen Überblick zu geben über Fragestellungen, Methoden, Qualität und Ergebnisse von RCTs, die von deutschen, universitären Abteilungen für Allgemeinmedizin durchgeführt wurden.

In der Vorgängerdissertation zu der vorliegenden Arbeit wurde bereits der *gesamte* Publikationsoutput der deutschen universitären Allgemeinmedizin für den Zeitraum 2000 bis 2010 untersucht [Grossmann, 2014, Grossmann et al., 2012, Schneider et al., 2012]. Hierbei wurden 794 Publikationen eingeschlossen, wobei sich ein starker Anstieg der Veröffentlichungen im zeitlichen Verlauf zeigte (22 Publikationen im Jahr 2000, 170 im Jahr 2010). Bei fast der Hälfte (49 %) der untersuchten Publikationen handelte es sich um Querschnittsstudien, gefolgt von Kohortenstudien (11 %), qualitativen Studien (10 %) und randomisierten Studien (8 %) [Grossmann et al., 2012].

Publikationen zu randomisierten Studien machen demnach nur einen kleinen Teil der durch die deutsche universitäre Allgemeinmedizin veröffentlichten Artikel aus, obwohl RCTs unbestritten den Goldstandard in der klinischen Forschung darstellen. Als Grundlage für Verbesserungen in diesem Bereich hat das DFG-Netzwerk „Klinische Studien in der Allgemeinmedizin“ [Gágyor et al., 2014] die Erstellung der vorliegenden Forschungsarbeit angeregt hat.

## 2 Methodik

### 2.1 Protokoll und Regelwerk

Einschlusskriterien, Suchmethoden, Selektionsprozess, Extraktion und Bewertung der Studienqualität wurden, soweit möglich, vorab in einem Protokoll festgelegt. Aufgrund der Heterogenität der untersuchten Studien musste jedoch eine Vielzahl von Einzelfallentscheidungen getroffen werden. War dies der Fall, so wurde für jede Einzelfallentscheidung in einem Regelwerk, das sukzessive durch den Autor der vorliegenden Arbeit erweitert wurde, das gewählte Vorgehen erläutert und begründet. Bei Unklarheiten wurde im weiteren Verlauf grundsätzlich primär das Regelwerk zu Hilfe genommen. Auf diesem Weg konnte sichergestellt werden, dass für mehrfach auftretende, gleichartige Probleme stets dasselbe Vorgehen gewählt wurde.

### 2.2 Einschlusskriterien

Studien, welche die nachfolgenden Einschlusskriterien 2.2.1 bis 2.2.3 erfüllten, wurden in der vorliegenden systematischen Übersichtsarbeit berücksichtigt. Diese Kriterien wurden anhand von Veröffentlichungen zu den in Frage kommenden Studien überprüft.

#### 2.2.1 Publikationszeitraum

Die Artikel aufzunehmender Studien mussten in dem Zeitraum 01.01.2000 bis 31.12.2014 (im Folgenden nur 2000 bis 2014 genannt) publiziert worden sein.

#### 2.2.2 Studiendesign

In die vorliegende Übersicht wurden ausschließlich randomisierte kontrollierte Studien, die im Folgenden mit „RCT“ (engl. randomised controlled trial) abgekürzt werden, aufgenommen. Abhängig von der Ebene, auf der die Randomisierung stattfindet, können dabei zwei Arten von RCTs unterschieden werden:

- **„klassische“ randomisierte Studien:**

Die Randomisierung erfolgt auf derselben Ebene, auf der auch die Zielvariablen bestimmt werden (z. B. werden Patienten randomisiert und an ebendiesem wird zu Studienende die Zielgröße, beispielsweise der Blutdruck, gemessen).

- **clusterrandomisierte Studien (im Folgenden auch „cRCTs“):**

Die Randomisierung erfolgt auf einer so genannten Cluster-Ebene (z. B. werden Praxen einer Schulungs- und einer Kontrollgruppe zugeteilt), die Zielvariablen werden aber auf einer anderen Ebene (z. B. bei den in der Praxis behandelten Patienten) gemessen [Campbell et al., 2012].

Sowohl „klassische“ als auch clusterrandomisierte Studien wurden in die vorliegende Arbeit aufgenommen. In welchem Stadium sich diese RCTs befanden, also ob sie noch in Planung, bereits angelaufen oder schon abgeschlossen waren, spielte für den Einschluss keine Rolle. In der späteren Analyse wurden allerdings RCTs, zu denen noch keine Ergebnisse publiziert worden waren (z. B. wenn nur ein Studienprotokoll vorhanden war), und RCTs, zu denen bereits Ergebnisse vorlagen, separat ausgewertet (3.2).

### **2.2.3 Affiliation zur deutschen universitären Allgemeinmedizin**

In der vorliegenden Arbeit sollten nur RCTs untersucht werden, die unter der Federführung oder Mitwirkung eines an einer deutschen Universität angesiedelten Instituts bzw. Lehrbereichs für Allgemeinmedizin durchgeführt worden waren. Diese Vorgabe wurde als erfüllt betrachtet, wenn es zu der in Frage kommenden Studie mindestens eine Publikation<sup>1</sup> gab, bei der eine Affiliation des Erst- und/oder Letztautors zu einem deutschen universitären Institut oder Lehrbereich für Allgemeinmedizin bestand. (Niedergelassene) Allgemeinmediziner *ohne* Anbindung an ein allgemeinmedizinisches Institut einer deutschen Universität erfüllten diese Vorgabe nicht. Bei Artikeln mit unvollständigen Angaben wurden die Informationen zur Affiliation der Autoren über andere Artikel bzw. dem Internet ergänzt.

## **2.3 Literatursuche**

Die Literatursuche musste entsprechend dem Einschlusskriterium 2.2.1 Artikel umfassen, die im Zeitraum 2000 bis 2014 publiziert worden waren.

Im Rahmen einer Vorgängerdissertation [Grossmann, 2014] war für den Zeitraum 01.01.2000 bis 31.12.2010 (im Folgenden nur 2000 bis 2010 genannt) bereits eine

---

<sup>1</sup> Hier bei musste es sich um eine „relevante“ Publikation handeln. Als „relevant“ wurde eine Publikation angesehen, wenn sie von den Ergebnissen der in Frage kommenden RCT berichtete oder wenn es sich um das Studienprotokoll einer entsprechenden RCT handelte. Eine Publikation wurde als „nicht relevant“ eingestuft, wenn sie z. B. „nur“ die Ergebnisse einer (ggf. parallel zu einer RCT durchgeführten) qualitativen Begleitstudie beinhaltete.

umfassende Literaturrecherche erfolgt, die alle für die vorliegende Arbeit relevanten Artikel aus dem Zeitraum 2000 bis 2010 enthielt (2.3.1).

Für den Zeitraum 01.01.2011 bis 31.12.2014 (im Folgenden nur 2011 bis 2014 genannt) basierte die Literatursuche auf zwei Säulen: Es erfolgte eine Online-Datenbanksuche sowie die Durchsicht von Publikationslisten (2.3.2). Als Kontrolle, ob sich mit der in der Online-Datenbank verwendeten Suchstrategie bzw. der Durchsicht von Publikationslisten alle relevanten Publikationen finden lassen würden, wurde der Suchzeitraum jeweils um ein Jahr verlängert (d. h. 2010 bis 2014 anstatt 2011 bis 2014). Somit wurde für das Jahr 2010 eine Überschneidung der Ergebnisse aus der Datenbanksuche im Internet bzw. den Publikationslisten und den aus der Vorgängerdissertation bereits bekannten Publikationen erzeugt. Anschließend wurde geprüft, ob alle aus der Vorgängerdissertation bekannten Publikationen aus dem Jahr 2010 auch in den Ergebnissen der Datenbanksuche bzw. in den Publikationslisten enthalten waren. Da dies der Fall war, konnte der in der Datenbanksuche verwendete Suchausdruck (2.3.2.1 und Tabelle 1) verifiziert werden.

### **2.3.1 2000 bis 2010: Publikationen aus Vorgängerdissertation**

Publikationen, die im Zeitraum 2000 bis 2010 veröffentlicht worden waren und von denen mindestens ein Autor mit einer allgemeinmedizinischen Abteilung einer deutschen Universität assoziiert gewesen war, waren bereits über die Literatursuche der Vorgängerdissertation dieser Arbeit identifiziert worden [Grossmann, 2014, die Ergebnisse dieser Vorgängerdissertation wurden zudem veröffentlicht: Grossmann et al., 2012, Schneider et al., 2012]. Die für die Literatursuche der Vorgängerdissertation verwendeten Suchstrategie entsprachen dabei weitestgehend derjenigen, die in der vorliegenden Arbeit für den Zeitraum 2010 bis 2014 angewendet wurde (2.3.2). Sie bestand also aus einer Online-Suche in der Datenbank Scopus und der Durchsicht von Publikationslisten (2.3.1.1 und 2.3.1.2).

Die Verfasserin der Vorgängerdissertation hatte alle Publikationen, welche durch ihre Literatursuche identifiziert worden waren, in die Kategorien „Interventionsstudien“, „diagnostische Studien“, „Beobachtungsstudien“ und „andere Studien“ eingeteilt. Da RCTs zu den Interventionsstudien zählen, mussten die für die vorliegende Arbeit relevanten Publikationen in der Kategorie „Interventionsstudien“ enthalten sein. Alle Artikel dieser Kategorie wurden deshalb im Volltext in das Literaturverwaltungsprogramm „EndNote“ übernommen.

Da in der Kategorie „Interventionsstudien“ neben randomisierten auch nichtrandomisierte Interventionsstudien enthalten waren und die Einschlusskriterien der vorliegenden Arbeit

hinsichtlich der Autorenschaft strenger waren als bei der Vorgängerdissertation<sup>1</sup>, mussten im späteren Selektionsprozess (2.4) einige Studien wieder ausgeschlossen werden.

### **2.3.1.1 Datenbanksuche Scopus (Vorgängerdissertation)**

Durch die Verfasserin der Vorgängerdissertation erfolgte die Scopus-Datenbanksuche am 20. und 21. Januar 2011 mittels folgendem Suchalgorithmus: „AFFIL (Allgemeinmedizin Stadt) OR AFFIL (general practice Stadt) OR AFFIL (family medicine Stadt) OR AFFIL (primary care Stadt)“. Die Scopussuche mit diesem Suchalgorithmus erfolgte mehrfach, wobei für jede Suche das Wort „Stadt“ durch einen anderen Städtenamen ersetzt wurde. Dabei wurde für all diejenigen deutschen Städte eine Suche durchgeführt, in denen sich zu diesem Zeitpunkt eine allgemeinmedizinische Abteilung oder ein allgemeinmedizinisches Institut befand. Bei Städtenamen, die einen Umlaut enthielten, wurde zusätzlich eine Suche mit „ae“, „oe“ oder „ue“ durchgeführt. [für Detailinformationen siehe Grossmann, 2014, S. 14 f.]

### **2.3.1.2 Durchsicht von Publikationslisten (Vorgängerdissertation)**

Neben dieser Datenbanksuche erfolgte die Durchsicht von Publikationslisten. Hierfür wurden durch die Verfasserin der Vorgängerdissertation im Oktober 2010 alle (zu diesem Zeitpunkt 32) allgemeinmedizinischen Abteilungen/Lehrbereiche in Deutschland angeschrieben und um möglichst vollständige Publikationslisten gebeten. Bei ausbleibender Antwort erfolgte eine telefonische Nachfrage. Durch die Publikationslisten konnten Arbeiten identifiziert werden, die durch die Scopussuche nicht erfasst worden waren. [Grossmann, 2014]

## **2.3.2 2010 bis 2014: Datenbanksuche Scopus und Durchsicht von Publikationslisten**

### **2.3.2.1 Datenbanksuche Scopus**

Die Datenbank Scopus ([www.scopus.com](http://www.scopus.com)) wurde genutzt, um Artikel zu finden, deren Publikationsdatum im Zeitraum 2010 bis 2014 lag. Diese seit 2004 bestehende Datenbank wurde vom Elsevier Verlag als weltgrößte Abstract- und Indizierungsdatenbank eingeführt

---

<sup>1</sup> In der Vorgängerdissertation musste ein Autor – unabhängig davon, ob es sich dabei um den Erst-, einen Mittel- oder den Letztautor handelte – eine Affiliation zu einer allgemeinmedizinischen Abteilung einer deutschen Universität vorweisen, damit das Einschlusskriterium der Autorenschaft erfüllt war. In der vorliegenden Arbeit hingegen musste diese Affiliation durch Erst- und/oder Letztautor erfüllt werden (2.2.3).

und enthält neben allen MEDLINE- und allen EMBASE-Dokumenten auch nichtenglischsprachige europäische Zeitschriften [Grossmann, 2014]. Folglich ist Scopus für eine Datenbanksuche im europäischen Raum besonders gut geeignet. Borgers [2008] bezeichnete Scopus als „ein auf Europa zugeschnittenes Nachweissystem für alle Wissenschaften“. Die erste Scopussuche erfolgte im Februar 2014, eine Aktualisierungssuche für Publikationen aus dem Jahr 2014 erfolgte im April 2015.

Um in der Suche gefunden zu werden, mussten Artikel – entsprechend der Einschlusskriterien (2.2) – folgende Bedingungen erfüllen:

- Ort: Deutschland bzw. Germany
- Publikationszeitraum: 01.01.2010–31.12.2014
- Studiendesign: Hinweise, dass es sich um eine RCT handelt
- Autor(en): Affiliation zur Allgemeinmedizin/general practice

Tabelle 1: Suchstrategie für Literatursuche Scopus

ID	Suche
#1	AFFILCOUNTRY(Deutschland)
#2	AFFILCOUNTRY(Germany)
#3	(#1 OR #2)
#4	AFFILORG(Allgemeinmedizin)
#5	AFFILORG(general practice)
#6	(#4 OR #5)
#7	DOCTYPE(ar)
#8	PUBYEAR > 2009
#9	PUBYEAR < 2015
#10	Random*
#11	Cluster
#12	(#10 OR #11)
#13	(#3 AND #6 AND #7 AND #8 AND #9 AND #12)

Der aus dieser Suchstrategie resultierende Suchausdruck für Scopus hatte folgendes Format:  
 (AFFILCOUNTRY(deutschland) OR AFFILCOUNTRY(germany) AND AFFILORG(Allgemeinmedizin) OR  
 AFFILORG(general practice)) AND DOCTYPE(ar) AND PUBYEAR > 2009 AND PUBYEAR < 2015 AND  
 (Random\* OR Cluster)

Anschließend wurde die Trefferliste aus Scopus exportiert und in die EndNote-Bibliothek übertragen. Dieser Export beinhaltete *alle* über die entsprechenden Publikationen in Scopus hinterlegten Informationen.

### 2.3.2.2 Durchsicht von Publikationslisten

Das Risiko, relevante Arbeiten zu übersehen, sollte möglichst gering gehalten werden. Deshalb wurde neben der Scopussuche eine zweite Literatursuche mit einer anderen Suchmodalität durchgeführt. Diese bestand aus der Durchsicht aller verfügbaren Publikationslisten deutscher allgemeinmedizinischer Universitätsinstitute.

Hierfür war es zunächst notwendig, die entsprechenden Institute zu identifizieren, was über die Website der Deutschen Gesellschaft für Allgemeinmedizin und Familienmedizin [DEGAM, 2013] erfolgte. Dabei fanden sich insgesamt 36 allgemeinmedizinische Abteilungen an deutschen Universitäten, deren Kontaktdaten zusammengestellt wurden.

Anfang April 2014 wurden mithilfe dieser Kontaktdaten 35 Allgemeinmedizininstitute (diese Zahl ergibt sich aus den 36 o. g. Abteilungen abzüglich des „eigenen“ Instituts) per Post kontaktiert und um Zusendung einer Liste von Publikationen (seit 2010), die mit an der jeweiligen Abteilung durchgeführten RCTs in Verbindung standen, gebeten. Alternativ konnte auch eine Gesamt-Publikationsliste versendet werden, auf der die entsprechenden Publikationen markiert waren. Auch für den Fall, dass eine Abteilung in dem o. g. Zeitraum *keine* RCTs durchgeführt hatte, wurde um eine kurze Rückmeldung gebeten. War bis Mitte Mai 2014 noch keine Antwort eingegangen, so wurde ein zweiter Brief mit einer Nachfrage bzw. Erinnerung verschickt. blieb auch der zweite Brief unbeantwortet, so wurde versucht, die Publikationslisten auf den Internetpräsenzen der jeweiligen Institute zu finden.

### 2.3.3 Durchsicht von Literaturangaben (engl. reference tracking)

Im Anschluss an die Literatursuche (2.3.1 und 2.3.2) erfolgte der Selektions- und Extraktionsprozess (2.4 und 2.5). Hierbei wurden *alle* in der Literatursuche gefundenen Publikationen gescreent (Selektion); im Selektionsprozess *nicht* aussortierte Artikel wurden vollständig gelesen, wobei die relevanten Studieninformationen schriftlich festgehalten wurden (Extraktion). Fanden sich in den Publikationen Verweise auf andere, für die vorliegende Arbeit potentiell relevante Publikationen/Studien, so wurden diese nachverfolgt und im Volltext besorgt (engl. reference tracking). Anschließend durchliefen diese nachträglich identifizierten Publikationen ebenfalls den regulären Selektionsprozess.

Auf diesem Weg konnten weitere Publikationen identifiziert werden, die unsere Einschlusskriterien erfüllten, jedoch nicht in der Literatursuche gefunden worden waren. Hierbei handelte es sich allerdings primär um Zusatzpublikationen zu bereits eingeschlossenen RCTs, weniger um Publikationen zu bis dahin unbekanntem RCTs.

Das hier beschriebene Vorgehen wird in Abbildung 3 und in Kapitel 2.5.1 näher erläutert.

## 2.4 Selektionsprozess

Um aus den durch die Literatursuche identifizierten Artikeln die irrelevanten Publikationen bzw. Studien aussortieren zu können, erfolgte der im Weiteren beschriebene Selektionsprozess. Die Selektion wurde durch den Autor dieser Arbeit mit Unterstützung des Betreuers bei Unklarheiten durchgeführt. Dabei wurde aus Gründen der Praktikabilität für die Selektion der Publikationen aus dem Zeitraum 2000 bis 2010, die bereits aus der Vorgängerdissertation bekannt waren (2.3.1), ein anderes Vorgehen gewählt (2.4.1) als für die Publikationen aus dem Zeitraum 2011 bis 2014, die mittels Scopussuche und Publikationslisten gefunden worden waren (2.3.2, 2.4.2).

### 2.4.1 2000 bis 2010: Selektion der Publikationen aus Vorgängerdissertation

Zunächst wurden die bereits aus der Vorgängerdissertation bekannten Publikationen selektiert, für die bereits der Volltext in gedruckter Form vorlag. Dabei wurde zuerst das Einschlusskriterium der Autorenschaft (2.2.3) geprüft. War dies erfüllt, so wurde in einem zweiten Schritt festgestellt, ob die entsprechende Publikation von einer RCT (2.2.2) berichtete (Abbildung 1). Dieses Vorgehen wurde gewählt, da die Beurteilung der Affiliation von Erst- und Letztautor nur sehr wenig Zeit in Anspruch nahm und somit die Selektion beschleunigt werden konnte.

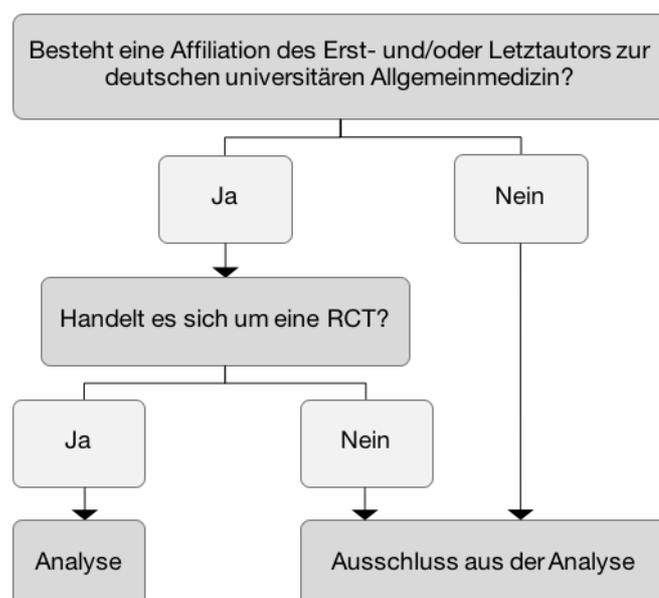


Abbildung 1: Selektion der Publikationen aus Vorgängerdissertationen (2000–2010)

## **2.4.2 2010 bis 2014: Selektion der Publikationen aus Datenbanksuche Scopus und Durchsicht von Publikationslisten**

Die Anwendung des in Abbildung 1 dargestellten Selektionsverfahrens war nur dann möglich, wenn die Volltexte der zu selektierenden Publikationen vorlagen, da die detaillierten Informationen zu den Affiliationen der Autoren nur auf den Volltexten zu finden sind. Im Gegensatz zu den über die Vorgängerdissertation identifizierten Publikationen lagen allerdings für die Studien, die mittels Datenbanksuche und Sichtung von Publikationslisten gefunden worden waren, *keine* Volltexte vor. Deshalb wurde für die Selektion dieser Publikationen ein anderes Vorgehen gewählt.

### **2.4.2.1 Selektion der Publikationen aus Datenbanksuche Scopus**

Für alle Publikationen, die über die Scopussuche erfasst worden waren, lagen lediglich Basisinformationen (wie Namen der Autoren, Publikationsdatum, Titel, Zeitschrift, Abstract etc.) vor. Diese Informationen waren verfügbar, da sie bei der Übertragung von Suchergebnissen einer Scopussuche in eine EndNote-Bibliothek automatisch gespeichert werden. Das in Abbildung 2 dargestellte Selektionsschema ermöglichte den Ausschluss irrelevanter Publikationen anhand dieser Basisinformationen (insbesondere der Abstracts), sodass bei diesen Publikationen auf die häufig zeit- und arbeitsintensive Volltextsuche verzichtet werden konnte. Nur für diejenigen Publikationen, die anhand der Basisinformationen *nicht* eindeutig ausgeschlossen werden konnten, erfolgte die Suche nach dem Volltext und die anschließende Prüfung desselben.

### **2.4.2.2 Selektion der mittels Durchsicht von Publikationslisten identifizierten Publikationen**

Von allen in den vorliegenden Publikationslisten (2.3.2.2) enthaltenen Publikationen, die *nicht* bereits durch die anderen Formen der Literatursuche (2.3.1 und 2.3.2.1) identifiziert worden waren, wurde auf PubMed ([www.pubmed.com](http://www.pubmed.com)) der Abstract gelesen. Der weitere Selektionsprozess entsprach dem in Abbildung 2 dargestellten. Über die Publikationslisten identifizierte Artikel wurden nur dann in die EndNote-Bibliothek aufgenommen, wenn für sie eine Volltextprüfung vorgenommen werden sollte (Abbildung 2).

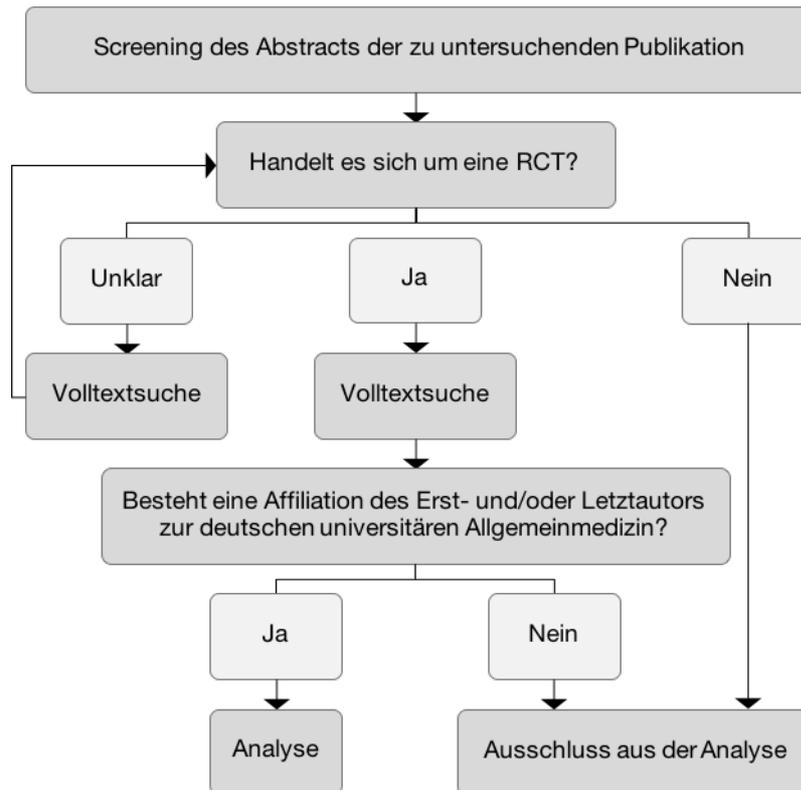


Abbildung 2: Selektion der Publikationen aus Datenbanksuche Scopus (2010–2014)

## 2.5 Extraktion von Studieninformationen

Vom Autor dieser Arbeit und dessen Betreuer wurde im Vorhinein festgelegt, welche Studieninformationen im Rahmen des Extraktionsprozesses erhoben werden sollten. Wie bereits in der Einleitung (1) erwähnt, wurden dabei auch Studieninformationen berücksichtigt, deren Erhebung das DFG-Netzwerk „Klinische Studien in der Allgemeinmedizin“ angeregt hatte. Anschließend wurde ein entsprechendes Extraktionsformular erstellt (Anhang A), mithilfe dessen die nachfolgend genannten Studieninformationen zu den eingeschlossenen RCTs standardisiert erhoben wurden. Im weiteren Verlauf erfolgte die Übertragung dieser Informationen in ein Excel-Datenblatt.

### 2.5.1 Extraktionsprozess – praktisches Vorgehen

Die Analyseeinheit der vorliegenden Arbeit war die einzelne RCT. Existierte zu einer RCT also nur *eine* Publikation, so genügte die Extraktion der Informationen aus ebendieser Publikation. Gab es hingegen mehrere Publikationen, die von ein- und derselben RCT berichteten (z. B. Studienprotokoll, Kurzzeit- und Langzeitergebnisse), so wurden zwar all

diese Publikationen in der Analyse berücksichtigt, die Studieninformationen aus diesen Publikationen aber dennoch auf *einem* Extraktionsformular zusammengeführt.

Für diese gemeinsame Extraktion aller Publikationen zu einer RCT war es allerdings nötig, alle Publikationen der entsprechenden RCT zu finden bzw. bereits vorliegende Publikationen als „zu *einer* RCT gehörend“ zu identifizieren („matching“). Dies gestaltete sich nicht immer einfach, da manche Autoren in ihren Publikationen zu einer RCT eine Bezugnahme auf andere Publikationen zu derselben RCT vermeiden. Um diesen Prozess zu vereinfachen, wurde das in Abbildung 3 dargestellte Vorgehen angewendet.

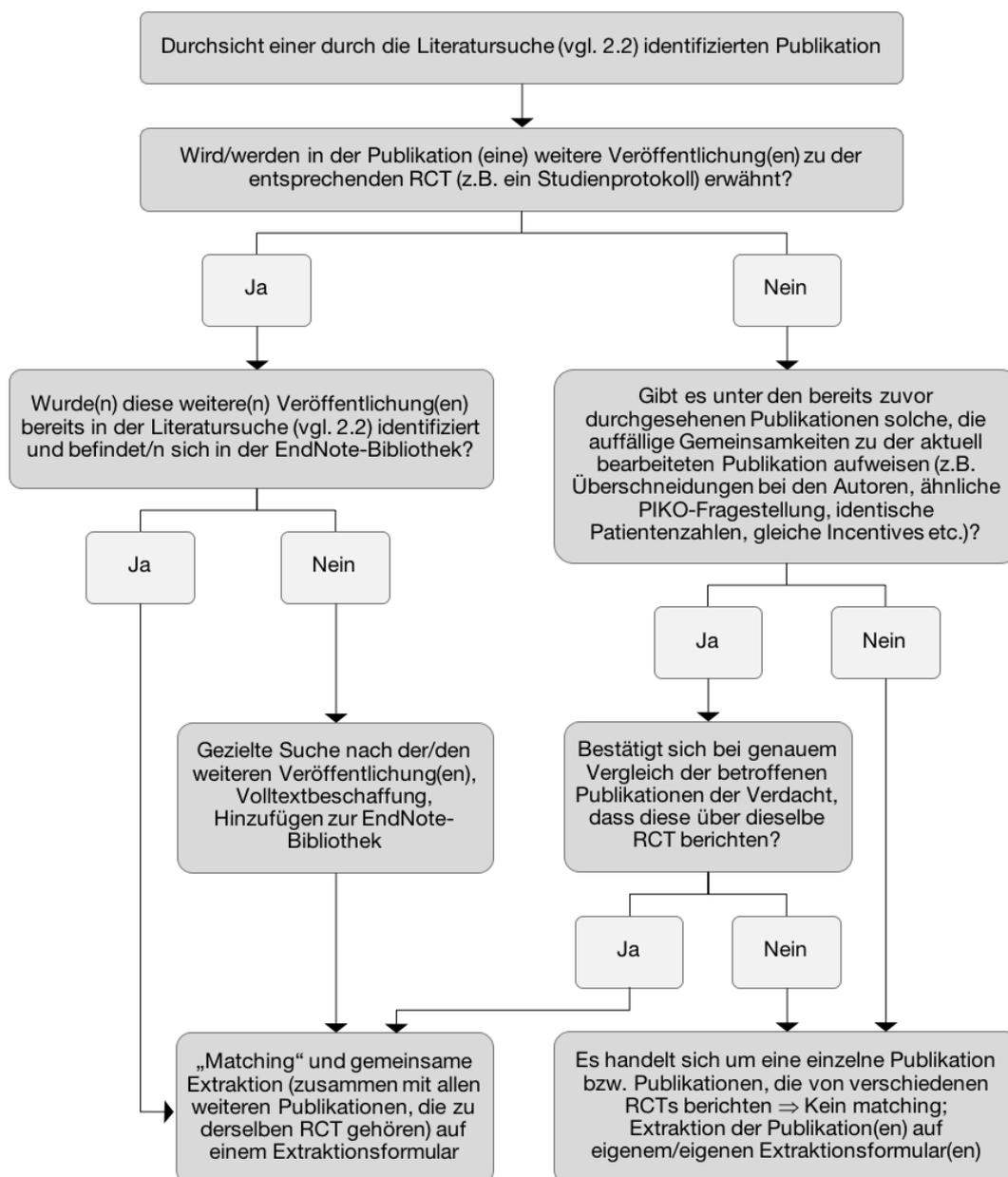


Abbildung 3: Praktisches Vorgehen zur Identifizierung zusammengehörender Artikel

Im weiteren Verlauf wurde in einer Excel-Tabelle eine Übersicht aus allen eingeschlossenen RCTs und den jeweils zugehörigen („gematchten“) Publikationen, die gemeinsam extrahiert werden sollten, erstellt.

Bei der gemeinsamen Extraktion von Studieninformationen zu einer RCT aus mehreren Publikationen wurde zunächst diejenige Publikation identifiziert, in der von den Hauptergebnissen der zu untersuchenden Studie berichtet wurde („Hauptpublikation“). Sie wurde durch den Autor dieser Arbeit vollständig gelesen, stellte den Schwerpunkt für die Extraktion dar und diente als wichtigste Quelle zum Ausfüllen des Extraktionsformulars. Trotzdem wurden auch alle Zusatzpublikationen gelesen und bei der Extraktion berücksichtigt.

### *Anmerkung:*

Neben den oben erwähnten „regulären“ Zusatzpublikationen, die – wie die jeweilige Hauptpublikation – direkt und unmittelbar von der zu analysierenden RCT berichteten (z. B. Studienprotokolle oder Follow-up Ergebnisse), existierten auch „indirekte“ Zusatzpublikationen. Dabei handelte es sich um Artikel, die nicht direkt von der entsprechenden RCT berichteten, sondern von kleineren „Zusatzstudien“, die zusammen mit der RCT durchgeführt worden waren bzw. auf in der RCT erhobene Daten zurückgriffen (z. B. Kosten-Nutzen-Analysen, Querschnittsstudien, qualitative Studien etc.). Lagen solche „indirekten“ Zusatzpublikationen vor und konnte das Extraktionsformular anhand der in Hauptpublikation und „regulären“ Zusatzpublikationen enthaltenen Informationen *nicht* vollständig ausgefüllt werden, so wurden die PDF-Dokumente der „indirekten“ Zusatzpublikationen mithilfe der sogenannten „Schlagwortliste“<sup>1</sup> (Anhang B) gezielt nach den fehlenden Informationen durchsucht (Beispiel 1). Auf diesem Weg konnten in einigen Fällen weitere Informationen gefunden werden.

Dennoch entschied sich der Autor dieser Arbeit nach Rücksprache mit dem Betreuer, die „indirekten“ Zusatzpublikationen in der Analyse nicht zu berücksichtigen. Der Grund für diese Entscheidung bestand in der übereinstimmenden Ansicht, dass die „indirekten“ Zusatzpublikationen die Einschlusskriterien im engeren Sinne nicht erfüllten. Dieser Vorgang sollte allerdings weniger als ein formaler Ausschluss der „indirekten“ Zusatzpublikationen gewertet, sondern vielmehr als eine Fokussierung auf die für die

---

<sup>1</sup> Bei der Schlagwortliste (Anhang B) handelt es sich um eine Liste von gängigen Schlagwörtern für bestimmte zu extrahierende Informationen auf Deutsch und Englisch. Die Schlagwortliste wurde durch den Autor dieser Arbeit während des Extraktionsprozesses sukzessive erweitert.

vorliegende Arbeit wesentlichen Artikel verstanden werden. Informationen, welche aus „indirekten“ Zusatzpublikationen gewonnen und in die Extraktionsformulare übernommen worden waren, verblieben dennoch in der weiteren Analyse (Beispiel 1).

In Abbildung 4 sowie in den beiden mit der vorliegenden Arbeit assoziierten Veröffentlichungen [Heinmüller et al., 2016a, Heinmüller et al., 2016b] wurde ebenfalls auf die indirekten Zusatzpublikationen verzichtet.

### Beispiel 1:

Sowohl in der Hauptpublikation [Becker et al., 2008] als auch in der „regulären“ Zusatzpublikation [Leonhardt et al., 2008] zu einer RCT waren keine Informationen zu Mitwirkungsanreizen (engl. incentives) für Probanden und/oder Ärzte zu finden. Deshalb wurden die PDF-Dokumente der vorliegenden „indirekten“ Zusatzpublikationen am Computer gezielt nach entsprechenden Schlagwörtern durchsucht, wie zum Beispiel:

- „Incentive“
- „Euro“ bzw. „€“
- „Aufwandsentschädigung“
- „Vergütung“

Mittels dieser Suche konnte in einer der „indirekten“ Zusatzpublikationen [Chenot et al., 2008] die Information gefunden werden, dass die an der Studie beteiligten Hausärzte eine Aufwandsentschädigung von 200 € erhalten hatten. Obwohl die „indirekten“ Zusatzpublikationen in der weiteren Analyse nicht berücksichtigt wurden, ging die Information über die Höhe Aufwandsentschädigung in die weitere Auswertung mit ein.

### 2.5.2 Formale Merkmale und Finanzierung

Folgende strukturelle Merkmale wurden erhoben:

- In welchem Jahr wurde der Artikel veröffentlicht?
- In welcher Sprache wurde die Publikation verfasst?
- Wie viele Autoren hatte die Publikation?
- Wie viele der Autoren hatten eine Affiliation zur Allgemeinmedizin?
- Hatte der Erstautor eine Affiliation zu einer allgemeinmedizinischen Abteilung einer deutschen Universität?

- Hatte der Letztautor eine Affiliation zu einer allgemeinmedizinischen Abteilung einer deutschen Universität?
- Welche deutschen universitären allgemeinmedizinischen Abteilungen waren an der Publikation beteiligt?
- Hatte einer der Autoren eine Affiliation zu einem Institut für medizinische Statistik oder Biometrie?
- War ein internationaler Autor an der Publikation beteiligt? D. h. hatte einer der Autoren eine Affiliation ins Ausland?
- Wie wurde die Studie finanziert?

### 2.5.3 Thematische Einordnung

Folgende Informationen wurden in Anlehnung an die in der evidenzbasierten Medizin weit verbreitete PICO-Frage (Patient, Intervention, Control, Outcome) [PICO, 2016] erhoben:

- Welche Informationen sind über die Patienten/Probanden bekannt, die an der vorliegenden Studie teilnahmen (Alter, Geschlecht, Vorerkrankungen etc.)?
- Welche Intervention(en) erhielten die Patienten/Probanden, die der Interventionsgruppe zugeteilt waren?
- Erhielten die Patienten/Probanden, die der Kontrollgruppe zugeordnet waren auch eine Intervention und wenn ja, welche?
- Welche Zielkriterien dienten zur Beurteilung des Interventionserfolgs (engl. outcome)?
- Können der in der vorliegenden Studie untersuchten Erkrankung Diagnosecodes nach ICPC-2 [ICPC-2, 2011] und ICD-10 [ICD-10, 2016] zugeordnet werden und falls ja, welche?

### 2.5.4 Studienbeteiligte und Studienteilnehmer

Da die Hausarztmedizin in Deutschland primär in Arztpraxen stattfindet, werden auch allgemeinmedizinische Studien häufig in den Praxen niedergelassener Ärzte durchgeführt. Deshalb wurden nicht nur zu den Patienten und Probanden, sondern auch zu den an den untersuchten Studien beteiligten Ärzten, deren Praxen und den beteiligten Teams, wie z. B. MFAs, Physiotherapeuten etc., Daten erhoben.<sup>1</sup>

---

<sup>1</sup> In der vorliegenden Arbeit werden Ärzte, Praxen und Assistententeams, die an den jeweiligen Studien mitgewirkt haben, als „Studienbeteiligte“ bezeichnet. Patienten bzw. Probanden werden hingegen „Studienteilnehmer“ genannt.

#### **2.5.4.1 Ärzte/Praxen/Teams („Studienbeteiligte“)**

Falls die in den Publikationen enthaltenen Informationen dies ermöglichten, wurden folgende Fragen beantwortet:

- Wie viele Ärzte bzw. wie viele Praxen erklärten sich initial bereit, sich an der Studie zu beteiligen?
- Wie viele Ärzte bzw. wie viele Praxen beteiligten sich zum Zeitpunkt der ersten Datenerhebung tatsächlich an der Studie (hatten also mindestens einen eingeschlossenen Patienten vorzuweisen)?
- Welche weiteren Informationen konnten über die an der Studie beteiligten Praxen noch in Erfahrung gebracht werden (handelte es sich z. B. um Lehrpraxen einer bestimmten Universität oder um Praxen aus einer bestimmten Region)?
- Erhielten die beteiligten Ärzte/Praxen/Teams (z. B. finanzielle) Anreize (engl. incentives)?
- Wie wurden die beteiligten Ärzte/Praxen/Teams rekrutiert?
- Wie groß war der Arbeitsaufwand von Ärzten/Teams für die Studie?

#### **2.5.4.2 Patienten/Probanden („Studienteilnehmer“)**

Auch in Bezug auf die an der Studie teilnehmenden Patienten bzw. Probanden wurden – falls möglich – Antworten auf die folgenden Fragen gefunden:

- Wie viele Patienten/Probanden waren zu Studienbeginn eingeschlossen?
- Wie viele Patienten/Probanden schlossen die Studie ab (engl. completer)?
- Wie viele Patienten/Probanden wurden in der statistischen Auswertung berücksichtigt?
- Erhielten Patienten/Probanden (z. B. finanzielle) Anreize für ihre Teilnahme an der Studie (engl. incentives)?
- Wie konnten die Patienten/Probanden für die Studienteilnahme gewonnen werden (Rekrutierung)?
- Wie viel Zeit mussten Patienten/Probanden für die Studienteilnahme investieren (Arbeitsaufwand)?

### 2.5.5 Details zum Studiendesign

In Bezug auf das Studiendesign wurden folgende Informationen erhoben:

- Handelte es sich bei der Studie um eine „klassische“ RCT oder um eine clusterrandomisierte RCT (cRCT)?
- Wie viele Gruppen (Studienarme) hatte die Studie?
- Wie lange dauerte die Studie für einen Teilnehmer (also die Zeit von Studieneinschluss bis -ende)?
- Wer/was wurde randomisiert? (D. h. was war die Randomisierungsebene (engl. unit of randomisation)?)
- Wer war verblindet (Patienten/Probanden? Ärzte/Behandler? „Outcomemessung“)?
- Auf welcher Ebene bzw. welchen Ebenen fand die Intervention statt (Arzt/Arzthelferin bzw. Praxisteam/Patient etc.)? Wer war das „Ziel“ der Intervention?
- „Kam“ die Intervention beim Patienten „an“? (D. h. hat der Patient etwas von der Intervention gemerkt?)
- Um welche Art von Intervention (z. B. Diagnostik, Therapie, Schulung etc.) handelte es sich?
- Wurde vor der Studie eine Fallzahlschätzung vorgenommen?
- Wurde eine Intention-to-treat (ITT-Analyse; d. h. Analyse aller eingeschlossenen Patienten) durchgeführt?
- Nur bei cRCTs (siehe auch Anmerkung am Ende des Absatzes):
  - Falls eine Fallzahlschätzung vorgenommen wurde, mit welchem geschätzten Intracluster-Korrelationskoeffizient (engl. intracluster correlation coefficient, ICC) erfolgte diese?
  - Wurde ein aus den tatsächlich erhobenen Daten berechneter ICC berichtet und falls ja, wie hoch war er?
  - Berücksichtigte die Analyse die Clusterstruktur der Studie?
- Wer hat die Daten erhoben, anhand derer die Zielkriterien beurteilt wurden?

#### *Anmerkung:*

Bei clusterrandomisierten Studien muss beachtet werden, dass die bei den einzelnen Elementen (z. B. Patienten) eines Clusters (z. B. einer Praxis) erhobenen Daten zu einem gewissen Grad in Zusammenhang stehen (d. h. die Stichprobe ist *nicht* gänzlich

unabhängig). Dies hat zur Folge, dass die effektive Fallzahl kleiner ist als die tatsächliche Anzahl an Studienteilnehmern, was wiederum durch größere Fallzahlen ausgeglichen werden kann. In welchem Maße die Fallzahlen hierfür erhöht werden ist von verschiedenen Faktoren abhängig. Hierzu zählen die durchschnittliche Clustergröße und das Ausmaß der Übereinstimmungen zwischen den Elementen eines Clusters. Dieses Übereinstimmungsmaß wird durch den sogenannten Intracluster-Korrelationskoeffizienten (engl. intracluster correlation coefficient, ICC) beschrieben. Da der ICC einer cRCT allerdings erst durch die Datenauswertung am Studienende berechnet werden kann, muss für die Fallzahlschätzung vor Studienbeginn ein geschätzter ICC herangezogen werden. In der Regel wird dabei ein ICC verwendet, der für andere, möglichst ähnliche cRCTs berichtet wurde. Des Weiteren muss die „Cluster-Struktur“ bei der statistischen Auswertung berücksichtigt werden. [Campbell et al., 2012]

### **2.5.6 Ergebnisse der untersuchten Studien**

Die Ergebnisse der untersuchten Studien, inklusive statistischer Kennzahlen, wurden auf dem Extraktionsformular in Textform festgehalten. Sie dienten als Grundlage für die spätere Zusammenfassung und Beurteilung der Studienergebnisse (0).

## **2.6 Beurteilung von Verzerrungsrisiken**

Selbst beim Vergleich zweier völlig identischer Interventionen (z. B. identischer Placebopräparate) im Rahmen einer RCT können aufgrund von Verzerrungen (engl. bias), die auf den ersten Blick sogar harmlos erscheinen mögen, signifikante Gruppenunterschiede hervorgerufen werden. Für die Praxis ist dies von Bedeutung, da sich nach demselben Muster in einer RCT natürlich auch ein völlig wirkloses Medikament im Vergleich zu einem Placebopräparat als signifikant überlegen darstellen kann. Die Beurteilung der Anfälligkeit einer RCT für solche Verzerrungen ist deshalb von großer Bedeutung, um die Aussagekraft der Studienergebnisse beurteilen zu können (d. h. um wissen zu können, wie viel Glauben man den Ergebnissen einer RCT schenken kann).

In der vorliegenden Arbeit wurde dieses Verzerrungsrisiko<sup>1</sup> anhand des „Risk of bias assessment tools“ der Cochrane Collaboration [Higgins et al., 2011] beurteilt. Dazu prüfte der Autor dieser Arbeit für jede eingeschlossene Studie die Kriterien 2.6.1 bis 2.6.6.

---

<sup>1</sup> Aus naheliegenden Gründen war die vollständige Beurteilung der Verzerrungsrisiken nur für bereits abgeschlossene RCTs mit Ergebnissen möglich, nicht jedoch für Studienprotokolle ohne Ergebnisse (3.3).

Anschließend erfolgte die Beurteilung des Verzerrungsrisikos, indem für jedes Kriterium eine der drei folgenden Risikokategorien vergeben wurde:

- „low risk of bias“:                   geringes Verzerrungsrisiko
- „high risk of bias“:                   hohes Verzerrungsrisiko
- „unclear risk of bias“:               mittleres (also weder geringes noch hohes)  
oder nicht beurteilbares Verzerrungsrisiko

In den Tabellen der Kapitel 2.6.1 bis 2.6.6 werden Beispiele für „hohe“ und „geringe“ Verzerrungsrisiken genannt. Die Bewertung „unclear risk of bias“ wurde in der Regel dann vergeben, wenn weder eine Bewertung mit „gering“ noch mit „hoch“ möglich war. Deshalb sind hierfür in den Tabellen häufig keine eigenständigen Beispiele angegeben.

Das vollständige Bewertungsinstrument ist im „Cochrane Handbook for Systematic Reviews of Interventions“ [Higgins et al., 2011] enthalten.

### 2.6.1 „Random sequence generation“

Für die Zuordnung von Patienten/Probanden in Interventions- und Kontrollgruppe muss vor der Gruppenzuteilung eine sog. Randomisierungssequenz (engl. random sequence) erstellt werden. Hierbei handelt es sich um einen „Dechiffrierungsschlüssel“, mithilfe dessen die Gruppenzugehörigkeit für jeden Patienten/Probanden ermittelt werden kann (Tabelle 3 kann hierfür als Beispiel dienen).

Im Rahmen einer RCT ist es zur Verhinderung einer Stichprobenverzerrung (engl. selection bias) wichtig, dass die Gruppenzuordnung nach dem Zufallsprinzip erfolgt. Folglich ist es auch wichtig, dass der o. g. Randomisierungssequenz eine auf dem Zufall basierende Systematik zugrunde liegt. Um zu prüfen, ob dies tatsächlich der Fall war, wird bei der Beurteilung des Kriteriums 2.6.1 genau untersucht, wie die Randomisierungssequenz generiert wurde (engl. random sequence generation). Tabelle 2 zeigt Positiv- und Negativbeispiele hierfür auf.

Tabelle 2: „Random sequence generation“ – Beurteilung des Verzerrungsrisikos

Verzerrungsrisiko	Die Randomisierungssequenz wurde erstellt mittels/nach...
gering	<ul style="list-style-type: none"> <li>- Zufallsgenerator am Computer.</li> <li>- Münzwurf.</li> <li>- Würfeln.</li> </ul>
hoch	<ul style="list-style-type: none"> <li>- geradem/ungeradem Geburtsdatum.</li> <li>- dem Datum der Aufnahme ins Krankenhaus.</li> <li>- persönlicher Präferenz des Arztes/Patienten.</li> </ul>

### 2.6.2 „Allocation concealment“

Eine Stichprobenverzerrung (engl. selection bias) kann, neben einer nicht auf dem Zufallsprinzip basierenden Randomisierungssequenz (2.6.1), auch durch eine mangelnde „Geheimhaltung“<sup>1</sup> (engl. allocation concealment) derselben hervorgerufen werden. Wissen die Patienten/Probanden nämlich im Vorhinein, ob sie im Falle einer Studienteilnahme der Interventions- oder der Kontrollgruppe zugeordnet werden würden, so könnte dieses Wissen (beispielsweise weil die Einteilung eines Patienten/Probanden in die Kontrollgruppe vorgesehen ist, der Patient/Proband aber in den „Genuss“ der zu untersuchenden Intervention kommen möchte und deshalb eine Zuordnung in die Interventionsgruppe präferiert) ihre Bereitschaft zur Studienteilnahme beeinflussen.

Eine Stichprobenverzerrung kann aber auch dann möglich sein, wenn Patienten/Probanden die für sie vorgesehene Gruppenzuordnung im Vorhinein *nicht* kennen. Dies wäre dann der Fall, wenn für die Patienten/Probanden nach der Bekanntgabe ihrer Gruppenzuordnung die Möglichkeit eines nachträglichen Gruppenwechsels gegeben wäre.

Sowohl die Vorhersehbarkeit der Gruppenzuordnung als auch die nachträgliche Einflussnahme auf die Gruppenzuordnung müssen natürlich nicht nur für die Patienten/Probanden, sondern auch für die an einer randomisierten kontrollierten Studie beteiligten Ärzte bzw. andere an der Studie beteiligten Forschungsmitarbeiter ausgeschlossen sein.

#### Beispiel 2:

Im Rahmen einer randomisierten kontrollierten Studie soll ein neues Präparat zur Behandlung von Asthma bronchiale gegen eine Placebokontrolle getestet werden. Ein an dieser Studie beteiligter Allgemeinmediziner ist angewiesen, alle sich in seiner Praxis vorstellenden Patienten mit Asthma bronchiale, welche die Einschlusskriterien erfüllen und sich zur Studienteilnahme bereit erklären, in die Studie aufzunehmen. Von der Studienzentrale hat der Arzt eine Zuordnungstabelle (Tabelle 3) erhalten, entsprechend derer er die eingeschlossenen Patienten der Interventionsgruppe bzw. der Kontrollgruppe zuordnen soll. (Die Tabelle wurde anhand eines computerbasierten Zufallsgenerators erstellt, sodass in Bezug auf 2.6.1 ein geringes Verzerrungsrisiko vorliegt.)

---

<sup>1</sup> „Allocation concealment“ kann auch als „verdeckte Zuteilung“ übersetzt werden [IQWiG, 2016].

Tabelle 3: Zuteilung der Asthmapatienten in Interventions- bzw. Kontrollgruppe

Asthmapatient 1	Intervention (neues Asthmamedikament)
Asthmapatient 2	Kontrolle (Placebo)
Asthmapatient 3	Kontrolle (Placebo)
Asthmapatient 4	Kontrolle (Placebo)
Asthmapatient 5	Intervention (neues Asthmamedikament)
Asthmapatient 6	Kontrolle (Placebo)
Asthmapatient 7	Intervention (neues Asthmamedikament)
Asthmapatient 8	Intervention (neues Asthmamedikament)

Diese Tabelle liegt dem Arzt vor; er weiß also, welcher Gruppe der nächste Asthmapatient zugeordnet werden müsste. Ist der Arzt jedoch aufgrund seiner klinischen Erfahrung davon überzeugt, dass z. B. für den nächsten Patienten, der laut Liste der Kontrollgruppe zugeteilt werden müsste, eine Behandlung mit einem Asthmamedikament sinnvoll sei, so besteht die Möglichkeit, dass der Arzt diesem Patienten von der Studienteilnahme abrät und ihn unabhängig von der Studie mit dem neuen Asthmamedikament therapiert.

Auch ein Wechsel der Reihenfolge wäre denkbar. Das heißt, dass der Arzt – entgegen den Anweisungen – z. B. Asthmapatient 1 der Kontrollgruppe und Asthmapatient 2 der Interventionsgruppe zuordnet (Tabelle 3).

Somit wäre für dieses Beispiel die Möglichkeit der Stichprobenverzerrung (engl. selection bias) gegeben und es müsste in Bezug auf das Item 2.6.2 („allocation concealment“) ein hohes Verzerrungsrisiko angenommen werden.

Das in diesem Beispiel genannte Problem ließe sich lösen, indem die o. g. Tabelle (Tabelle 3) in einer Randomisierungszentrale hinterlegt und „geheim gehalten“ würde. Der Arzt könnte somit nicht über die geplante Gruppenzuordnung für den nächsten einzuschließenden Studienpatienten Bescheid wissen. Er müsste die Randomisationszentrale anrufen, sobald ein Patient die Einschlusskriterien für die Studie erfüllt und sich zur Teilnahme bereit erklärt, um von ihr die Gruppenzuordnung dieses Patienten mitgeteilt zu bekommen. Auf diesem Weg wäre eine Beeinflussung der Gruppenzuordnung durch Arzt oder Patient unmöglich.

Tabelle 4 zeigt Beispiele für die Bewertung des Verzerrungsrisikos in Bezug auf das „allocation concealment“.

Tabelle 4: „Allocation concealment“ – Beurteilung des Verzerrungsrisikos

Verzerrungsrisiko	Die Patienten/Probanden und das Studienteam, das an der Aufnahme der Patienten/Probanden in die Studie beteiligt war, konnten die geplante Gruppenzuordnung eines Patienten/Probanden im Vorhinein...
gering	<ul style="list-style-type: none"> <li>- nicht kennen, weil es sich um eine zentralisierte Gruppenzuteilung handelte (d. h. mittels Telefon, Internet etc.).</li> <li>- nicht kennen, weil sich die Informationen über die Gruppenzuordnungen der Probanden auf Zetteln in undurchsichtigen, versiegelten und durchnummerierten Briefkuverts befanden.</li> <li>- nicht kennen, weil die Behältnisse, die beispielsweise Medikament oder Placebo enthielten, alle gleich aussahen und durchnummeriert waren.</li> </ul>
hoch	<ul style="list-style-type: none"> <li>- evtl. kennen, weil sich die Informationen über die Gruppenzuordnungen der Probanden auf Zetteln in Briefkuverts befanden, die nicht versiegelt waren oder durchsichtig waren oder ähnliches.</li> <li>- evtl. kennen, weil die Probanden abwechselnd der Interventions- und der Kontrollgruppe zugeteilt wurden.</li> <li>- evtl. kennen, weil die Zuteilung in die Gruppen anhand des Geburtsdatums der Probanden stattfand und somit bei Kenntnis der Zuteilungsregel und des Geburtsdatums des Probanden die Gruppenzuteilung vorhersehbar war.</li> <li>- kennen, weil es sich um eine clusterrandomisierte Studie handelte (d. h., dass z. B. Arztpraxen in Interventions- oder Kontrollpraxen randomisiert wurden und alle Probanden einer Praxis entweder die Intervention oder die Kontrollintervention erhielten) und das Studienteam bei der Rekrutierung der Probanden bereits wusste, ob es sich bei der Praxis um eine Interventions- oder eine Kontrollpraxis handelte.</li> </ul>

### 2.6.3 „Blinding of participants and personnel“

Auch wenn Patienten/Probanden und/oder Ärzte/Studienassistenten die Gruppenzugehörigkeit der Patienten/Probanden erst *nach*<sup>1</sup> der finalen, nicht mehr abänderbaren Gruppenzuteilung erfahren, können Verzerrungen entstehen. Grund dafür ist, dass dieses Wissen bei den Studienärzten zu einer (meist unbeabsichtigten) Ungleichbehandlung der Patienten/Probanden (engl. performance bias) führen kann (abgesehen natürlich von den Unterschieden in der Behandlung der Patienten/Probanden aufgrund der verschiedenen Studieninterventionen). Bei den Patienten/Probanden kann das Wissens um deren Gruppenzugehörigkeit zu unterschiedlichen Erwartungseffekten führen. Deshalb sollte eine Verblindung von Studienteilnehmern *und* Studienpersonal (engl. blinding of participants and personnel) angestrebt werden („doppelte Verblindung“).

---

<sup>1</sup> Dieses Wissen um die *tatsächliche* Gruppenzuteilung ist nicht zu verwechseln mit dem Wissen um die *geplante* Gruppenzuteilung vor Studienbeginn (2.6.1).

**Beispiel 3:**

Im Rahmen einer RCT soll die Wirkung eines speziellen osteopathischen Verfahrens („strain-counterstrain“) mit einer „sham“-Behandlung<sup>1</sup> bei Patienten mit einer akuten Episode von Nackenschmerzen mit Bewegungseinschränkungen in der Halswirbelsäule (HWS) untersucht werden [Klein et al., 2013]. Die Randomisierungssequenz wird mittels eines Zufallsgenerators am Computer erstellt und auch die Informationen über die Gruppenzuordnung der Patienten werden geheim gehalten (d. h. nach 2.6.1 und 2.6.2 besteht nur ein geringes Verzerrungsrisiko). Allerdings ist aufgrund der Art der untersuchten Intervention keine Verblindung von Ärzten/Patienten möglich (d. h. Arzt und ggf. auch Patient kennen die Gruppenzuordnung der Patienten). Die „Outcomemessung“ erfolgt durch einen verblindeten Untersucher, der die Gruppenzuordnung der Patienten nicht kennt, wobei der Bewegungsumfang der HWS nach der Intervention im Vergleich zum Ausgangswert als Zielkriterium gilt.

Die mangelnde Verblindung von Arzt und Patienten stellt in diesem Fall keinen methodischen Mangel der Studie dar, da eine Verblindung aufgrund der Art der untersuchten Intervention nicht möglich war. Trotzdem muss deshalb für 2.6.3 ein hohes Verzerrungsrisiko angenommen werden, was die Aussagekraft der Studienergebnisse schmälert.

Beispiele für die Bewertung des Risikos einer Verzerrung von Studienergebnissen in Bezug auf die Verblindung von Probanden und Studienpersonal sind in Tabelle 5 aufgelistet.

Tabelle 5: „Blinding of participants and personnel“ – Beurteilung des Verzerrungsrisikos

Verzerrungsrisiko	Probanden und Studienpersonal waren...
gering	<ul style="list-style-type: none"> <li>- verblindet und es ist unwahrscheinlich, dass eine Entblindung stattgefunden hat.</li> <li>- nicht verblindet bzw. nur teilweise verblindet, aber die Studienergebnisse wurden durch diese mangelnde Verblindung wahrscheinlich nicht beeinflusst.</li> </ul>
hoch	<ul style="list-style-type: none"> <li>- verblindet, aber die Wahrscheinlichkeit, dass eine Entblindung mit nachfolgender Beeinflussung der Studienergebnisse stattgefunden hat, ist hoch.</li> <li>- nicht verblindet bzw. nur teilweise verblindet und diese mangelnde Verblindung hat wahrscheinlich das Studienergebnis beeinflusst.</li> </ul>

<sup>1</sup> Als „sham“-Behandlung bezeichnet man eine wirkungslose Behandlung, ähnlich einem Placebo.

### 2.6.4 „Blinding of outcome assessment“

Vor der Durchführung einer RCT wird in der Regel eine quantifizierbare Größe als Zielkriterium (engl. outcome<sup>1</sup>) festgelegt (z. B. Höhe des Blutdrucks, Punkte in einem Fragebogen zur Lebensqualität o. ä.). Nach der Studie wird anhand dieses Kriteriums beurteilt, wie sich die Interventionsgruppe zur Kontrollgruppe verhalten hat. Dabei besteht die Möglichkeit der „Detektionsverzerrung“ (engl. detection bias), wenn die Erhebung der Zielgrößen unverblindet erfolgt und wenn diese mangelnde Verblindung einen Einfluss auf die Studienergebnisse haben könnte.

#### Beispiel 4:

In einer randomisierten Studie wird ein neues Lehrkonzept (= Intervention) für den Untersuchungskurs im Medizinstudium mit dem herkömmlichen Untersuchungskurs (= Kontrolle) verglichen [Blank et al., 2013]. Am Ende der Studie soll mittels einer sog. OSCE (engl. objective structured clinical examination), also einer speziellen mündlich-praktischen Prüfung, das Wissen und die Fertigkeiten der Studenten in Bezug auf die körperliche Untersuchung beurteilt werden. Die Leistung eines jeden Studenten wird dabei durch einen verblindeten Prüfer bewertet, sodass das Verzerrungsrisiko in Bezug auf die verblindete „Outcomemessung“ (2.6.4) als gering einzustufen ist.

Wären die Prüfer jedoch nicht verblindet gewesen, so hätte die Gefahr bestanden, dass unterbewusst – beispielsweise aufgrund der persönlichen Überzeugung eines Prüfers, dass das neue Lehrkonzept dem herkömmlichen Untersuchungskurs überlegen ist – den Prüflingen, die das neue Lehrkonzept durchlaufen haben, bessere Noten gegeben werden als den Prüflingen, die den herkömmlichen Kurs besucht haben (selbst wenn die Leistung der Prüflinge aus beiden Gruppen in Wirklichkeit gleich gut war). In diesem Fall könnte sich demnach fälschlicherweise eine Überlegenheit des neuen gegenüber des herkömmlichen Lehrkonzeptes zeigen. Wären die Prüfer also nicht verblindet gewesen, so hätte in Bezug auf 2.6.4 ein hohes Verzerrungsrisiko angenommen werden müssen.

Tabelle 6 zeigt anhand einiger Beispiele, wie die Bewertung des Risikos einer Verzerrung von Studienergebnissen in Bezug auf die verblindete Zielgrößenerhebung vorgenommen wurde.

---

<sup>1</sup> Meist wird zwischen primären und sekundären Zielkriterien unterschieden (engl. primary/secondary outcomes).

Tabelle 6: „Blinding of outcome assessment“ – Beurteilung des Verzerrungsrisikos

Verzerrungsrisiko	Die Studienbewerter (engl. outcome assessors)...
gering	<ul style="list-style-type: none"> <li>- waren nicht verblindet, aber es ist trotzdem unwahrscheinlich, dass die mangelnde Verblindung der Studienbewerter die Studienergebnisse beeinflusst hat.</li> <li>- waren verblindet und eine Entblindung ist unwahrscheinlich.</li> </ul>
hoch	<ul style="list-style-type: none"> <li>- waren nicht verblindet und die Wahrscheinlichkeit, dass die mangelnde Verblindung der Studienbewerter einen Einfluss auf die Studienergebnisse gehabt hat, ist groß.</li> <li>- waren verblindet, aber es ist davon auszugehen, dass eine Entblindung mit nachfolgender Beeinflussung der Studienergebnisse stattgefunden hat.</li> </ul>

### 2.6.5 „Incomplete outcome data“

Bei fast allen RCTs gibt es Patienten/Probanden, die ihre Studienteilnahme vorzeitig beenden, sodass für sie keine weitere Datenerhebung mehr möglich ist. Die daraus resultierenden „Lücken“ in den Studiendaten können die Studienergebnisse verzerren (engl. attrition bias). Davon ist allerdings nur auszugehen, wenn die Anzahl der Studienabbrecher sehr hoch ist, wenn sich die Anzahl der Studienabbrecher aus Interventions- und Kontrollgruppe stark unterscheidet oder wenn die Gründe für die Studienabbrüche in Interventions- und Kontrollgruppe unterschiedlich sind. Fand eine sogenannte Intention-to-treat-Analyse (ITT-Analyse) statt, so konnte ein größerer Anteil an fehlenden Daten toleriert werden, ohne dass sich hierdurch das Verzerrungsrisiko erhöhte.

#### Beispiel 5:

Unvollständige Daten zu den in einer RCT untersuchten Zielkriterien können verursacht werden durch...

- Probanden, die ihre Studienteilnahme vor der Datenerhebung beenden.
- unkooperative Studienteilnehmer, die z. B. kein Tagebuch führen oder nicht bereit sind einen Fragebogen auszufüllen.
- Studienteilnehmer, die vom Studienteam für die Datenerhebung eingeladen werden sollen, aber (z. B. aufgrund eines Umzugs) nicht mehr kontaktierbar sind (engl. lost to follow-up).

Um das Verzerrungsrisiko durch fehlende Daten beurteilen zu können, muss zunächst der Anteil derjenigen Studienteilnehmer bestimmt werden, die *nicht* an der (finalen) Datenerhebung teilgenommen haben:

$$\text{fehlende Daten (in \%)} = \frac{\text{eingeschlossene Studienteilnehmer} - \text{completer}^1}{\text{eingeschlossene Studienteilnehmer}}$$

Beispiel 6:

An einer RCT nehmen initial 100 Probanden (= eingeschlossene Studienteilnehmer) teil. Im Verlauf brechen jedoch 25 Probanden die Studienteilnahme ab, sodass nur 75 Probanden (= „completer“) für die (finale) Datenerhebung zur Verfügung stehen. Entsprechend der o. g. Formel würde der Anteil der fehlenden Daten in diesem Beispiel 25 % betragen.

Bei der Beurteilung des Verzerrungsrisikos wurde bei dem Punkt „Incomplete outcome data“ von den Vorschlägen des „Risk of bias assessment tools“ der Cochrane Collaboration [Higgins et al., 2011] abgewichen, da diese zu wenig eindeutig waren. Zusammen mit dem Betreuer wurde durch den Autor dieser Arbeit ein neuer Bewertungsschlüssel entworfen (Tabelle 7 und Anhang C).

Tabelle 7: „Incomplete outcome data“ – Beurteilung des Verzerrungsrisikos

Verzerrungsrisiko	Der Anteil der fehlenden Daten betrug...
gering	- ≤ 5 %.
	- ≤ 20 % <i>und</i> die Anzahl der Studienabbrecher war gleichmäßig unter den Gruppen verteilt <i>und</i> die Gründe für den vorzeitigen Studienabbruch waren in den Gruppen ähnlich <i>und</i> es wurde eine ITT-Analyse durchgeführt.
hoch	- > 20 % und es fand <i>keine</i> ITT-Analyse statt.
	- > 30 %.
	- Die Anzahl der Studienabbrecher in den verschiedenen Gruppen war sehr unterschiedlich.
unklar	Die Bewertung „unklar“ wird vergeben, wenn weder die Kriterien für ein geringes, noch die Kriterien für ein hohes Verzerrungsrisiko gegeben sind.

Abkürzungen (weitere Details zu den Abkürzungen finden sich im Abkürzungsverzeichnis):

ITT-Analyse = Intention-to-treat-Analyse

<sup>1</sup> Als „completer“ werden diejenigen Patienten/Probanden bezeichnet, die bis zum regulären Ende an der Studie teilnehmen und ihre Teilnahme *nicht* vorzeitig beenden.

### 2.6.6 „Selective reporting“

Auch durch eine selektive Berichterstattung seitens des Autors einer Publikation können Verzerrungen erzeugt werden. Von selektiver Berichterstattung (engl. selective reporting) spricht man, wenn beispielsweise wichtige Ergebnisse verschwiegen oder ungeeignete Analysemethoden verwendet werden, sodass die Studienergebnisse in ein falsches Licht gerückt werden. Selektive Berichterstattung hat demnach zur Folge, dass bei dem Leser der Publikation ein verzerrter Gesamteindruck über die Wirksamkeit der untersuchten Interventionen entsteht.

Zur Bewertung des Verzerrungsrisikos durch selektive Berichterstattung (engl. reporting bias) wurde anhand der nachfolgenden Tabelle 8 vorgegangen:

Tabelle 8: „Selective reporting“ – Beurteilung des Verzerrungsrisikos

Verzerrungsrisiko	Kriterien, die für die Vergabe des jeweiligen Verzerrungsrisikos erfüllt sein mussten
gering	<ul style="list-style-type: none"> <li>- Das Studienprotokoll ist verfügbar und alle dort angekündigten Haupt- und Nebenzielkriterien wurden in der Hauptpublikation entsprechend den Ankündigungen im Studienprotokoll genannt.</li> <li>- Das Studienprotokoll ist nicht verfügbar, aber alle Studienergebnisse, die bei einer Studie dieser Art von Interesse sind, wurden in der Publikation genannt.</li> </ul>
hoch	<ul style="list-style-type: none"> <li>- Nicht alle Ergebnisse der im Studienprotokoll angekündigten Zielkriterien wurden veröffentlicht.</li> <li>- Es gibt Zielkriterien, die mittels unüblicher Einheiten, Analysemethoden etc. angegeben wurden.</li> <li>- Die in der Hauptpublikation als Hauptzielkriterien deklarierten Ergebnisse, waren nicht im Studienprotokoll als solche angekündigt worden.</li> <li>- Ergebnisse von Zielkriterien werden nur unvollständig berichtet.</li> <li>- In der Publikation wird ein wichtiges Zielkriterium, das bei einer Studie dieser Art zu erwarten wäre, nicht erwähnt.</li> </ul>

## 2.7 Zusammenfassung und Beurteilung von Studienergebnissen

Studienergebnisse waren nur für bereits abgeschlossenen Arbeiten vorhanden (33 Studien), nicht hingegen für geplante Studien, zu denen nur Studienprotokolle vorlagen (17 Studien). Somit bezieht sich dieses Kapitel nur auf die 33 Studien, zu denen Ergebnisse vorlagen.

### 2.7.1 „Vote count“

Um eine möglichst einfache und zusammenfassende Beurteilung der zuvor erhobenen Studienergebnisse (2.5.6) zu ermöglichen, wurde jede untersuchte Studie einer Bewertungskategorie [ - / (-) / 0 / (+) / + ] in einem so genannten „vote count“ [Hedges et al., 1980, Melchart et al., 1999] zugeordnet (Tabelle 9).

Tabelle 9: Beurteilungsgrundlage für „vote count“

Bewertungskategorie	Bedeutung
+	Die Interventionsgruppe ist der Kontrollgruppe in Bezug auf das Zielkriterium signifikant überlegen.
(+)	Eine Überlegenheit der Interventionsgruppe gegenüber der Kontrollgruppe in Bezug auf das Zielkriterium ist erkennbar, aber ohne statistische Signifikanz.
0	Keine der beiden Gruppen ist der anderen in Bezug auf das Zielkriterium überlegen.
(-)	Eine Überlegenheit der Kontrollgruppe gegenüber der Interventionsgruppe in Bezug auf das Zielkriterium ist erkennbar, aber ohne statistische Signifikanz.
-	Die Kontrollgruppe ist der Interventionsgruppe in Bezug auf das Zielkriterium signifikant überlegen.

Die Zuordnung einer Studie zu einer einzigen Bewertungskategorie gestaltete sich allerdings nur dann problemlos, wenn in der Studie nur *ein* Zielkriterium zu genau *einem* Zeitpunkt erhoben wurde. Dies war jedoch nur in den wenigsten Studien der Fall.

Bei den meisten Studien erfolgte eine Untersuchung mehrerer Zielkriterien zu mehreren Zeitpunkten, sodass die Möglichkeit bestand, dass jedes Zielkriterium zu jedem Erhebungszeitpunkt einer anderen Bewertungskategorie zugeordnet werden musste. Somit war es schwierig, der gesamten Studie eine einzige zusammenfassende Bewertungskategorie zuzuteilen, was wiederum die Gesamtbeurteilung der Sinnhaftigkeit der jeweiligen Intervention erschwerte.

Um in solchen Fällen dennoch eine Gesamtbeurteilung der Studie, also die Zuordnung der gesamten Studie zu einer Bewertungskategorie, zu ermöglichen, wurden durch den Autor dieser Arbeit in Absprache mit dem Betreuer Einzelfallentscheidungen getroffen. Dabei wurde versucht, alle Zielkriterien zu berücksichtigen, das Hauptzielkriterium allerdings stärker zu gewichten.

Bei der Bewertung der Studien fiel häufig auf, dass die Autoren der Publikationen in ihren Schlussfolgerungen die Studienergebnisse teilweise anders beurteilten und interpretierten als der Autor dieser Arbeit. Um diese verschiedenen Interpretationen der Studienergebnisse abbilden zu können, wurde eine zweite Bewertungskategorie eingeführt, die „author’s conclusion“, also die Bewertung der Studienergebnisse durch den Autor der jeweiligen untersuchten Studie (Benotung ebenfalls nach Tabelle 9). Die weiter oben beschriebene Bewertung durch den Autor der vorliegenden Arbeit hingegen wird im Folgenden „reviewer’s conclusion“ genannt.

### 2.7.2 Metaanalyse

Im Rahmen der statistischen Auswertung einer randomisierten Studie wird die Wahrscheinlichkeit, mit welcher die Unterschiede zwischen den verschiedenen Gruppen in Bezug auf das Zielkriterium auf den Zufall zurückzuführen sind, errechnet. Nur wenn diese Wahrscheinlichkeit (auch p-Wert genannt) einen bestimmten Schwellenwert nicht überschreitet, wird davon ausgegangen, dass die Gruppenunterschiede der untersuchten Intervention geschuldet sind (statistische Signifikanz). Der Schwellenwert (Signifikanzniveau) wird dabei häufig auf 5 % festgelegt (p-Wert  $\leq 5\%$ ).

Hierbei bleiben allerdings andere Faktoren, die für die Beurteilung der Sinnhaftigkeit einer Intervention ebenfalls wichtig wären, unberücksichtigt. Zum Beispiel kann bei einer randomisierten Studie, in der sich Interventions- und Kontrollgruppe in Bezug auf das Zielkriterium nur *minimal* unterscheiden, allein aufgrund einer sehr großen Stichprobe ein signifikantes Ergebnis erzielt werden. Aus diesem Grund kann zur besseren Beurteilbarkeit eine andere statistische Größe herangezogen werden, die sogenannte Effektstärke (engl. effect size). Mithilfe der Effektstärken ist es möglich (unter Berücksichtigung anderer Faktoren, wie z. B. der Stichprobengröße) die Gewichtigkeit der Gruppenunterschiede zu beurteilen.

Die Abschätzung der Effektgrößen der in der vorliegenden Arbeit untersuchten Studien erfolgte jeweils in Bezug auf das Hauptzielkriterium oder – falls kein eindeutiges Hauptzielkriterium vordefiniert war – auf dasjenige Zielkriterium, das durch Autor und Betreuer der vorliegenden Arbeit als relevantestes Zielkriterium eingestuft wurde. Hierfür wurden die Rohdaten in die Kalkulationstabelle des im Internet verfügbaren Programms „Comprehensive Meta-Analysis Software“ ([www.meta-analysis.com](http://www.meta-analysis.com)) eingegeben, welches vier verschiedene Kombinationen aus statistischen Größen verwerten kann:

- Mittelwerte, Standardabweichungen und Stichprobengrößen
- mittlere Differenzen (engl. mean differences), Stichprobengrößen und p-Werte bzw. Konfidenzintervalle
- Ereignisse (engl. events) und Anzahl an Studienteilnehmern pro Gruppe
- Chancenverhältnis (engl. odds ratio) und Konfidenzintervalle

Das Rechenprogramm ermöglichte die Konversion dieser verschiedenen Arten statistischer Daten in standardisierte mittlere Differenzen (engl. standardised mean differences, SMDs) mit 95 %-Konfidenzintervallen. Positive Werte sprachen dabei für eine Überlegenheit der Interventionsgruppe. In der vorliegenden Arbeit wurden SMDs in Anlehnung an Schünemann et al. [2011] folgendermaßen eingeteilt:

- $SMD \leq 0,4$                     kleiner Effekt
- $0,4 < SMD \leq 0,7$             moderater Effekt
- $0,7 < SMD$                     großer Effekt

Es sei darauf hingewiesen, dass die Metaanalyse durch den Betreuer der vorliegenden Dissertation durchgeführt wurde. Freundlicherweise stellte er die Ergebnisse für die Verwendung in dieser Arbeit zur Verfügung.

## **3 Ergebnisse**

### **3.1 Literatursuche und Selektionsprozess**

Insgesamt lieferten die verschiedenen Suchstrategien 639 Textreferenzen. Wie sich diese im Detail zusammensetzten wird im Folgenden sowie in Abbildung 4 erläutert.

#### **3.1.1 Publikationen aus Vorgängerdissertation, Scopussuche, „reference tracking“**

Die 72 Textreferenzen, welche durch die Verfasserin der Vorgängerdissertation als „Interventionsstudien“ klassifiziert worden waren [Grossmann, 2014, Abb. 3], wurden zur weiteren Prüfung der Einschlusskriterien in die vorliegende Arbeit übernommen. Über die Scopussuche für den Zeitraum 2010 bis 2014 wurden insgesamt 524 Textreferenzen identifiziert, mittels „reference tracking“ fanden sich 42 weitere Publikationen.

#### **3.1.2 Publikationslisten**

Von 27 der 35 angeschriebenen Abteilungen (2.3.2.2) erhielt das Institut für Allgemeinmedizin der TU München eine Antwort. Da für weitere 4 der 8 verbleibenden Institute auf der jeweiligen Homepage eine aktuelle Publikationsliste heruntergeladen werden konnte, lagen für insgesamt 32 von 36 (inkl. TU München) deutschen universitären Abteilungen für Allgemeinmedizin entsprechende Informationen vor. 14 der 32 o. g. Institute gaben an, seit 2010 keine RCTs durchgeführt zu haben und verzichteten deshalb auf die Zusendung einer Publikationsliste.

Für die 18 verbleibenden Institute lagen Publikationslisten vor, welche im weiteren Verlauf mit der EndNote Datenbank abgeglichen wurden. Publikationen, die nicht bereits in der EndNote-Bibliothek enthalten waren, gingen in den Selektionsprozess ein (2.4.2.2 und Abbildung 2).

Auf diesem Weg konnte nur 1 Publikation, die unsere Einschlusskriterien erfüllte und nicht ebenfalls über eine der anderen Literatursuchen gefunden worden war, identifiziert werden.

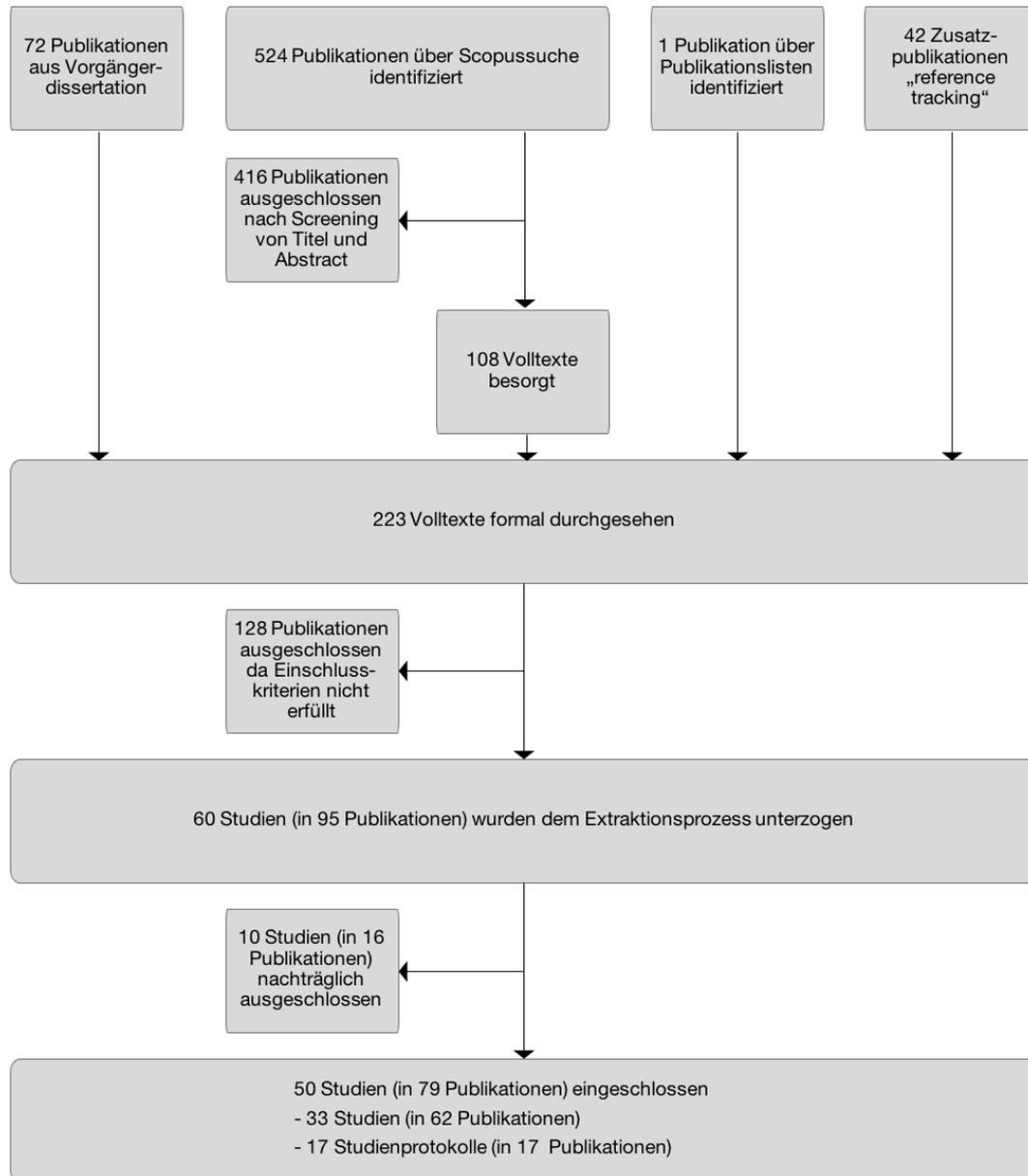


Abbildung 4: Flowchart zur schematischen Darstellung des Selektionsprozesses

#### 3.1.3 Selektionsprozess

Von den 524 über die Scopussuche gefundenen Publikationen konnten 416 bereits im ersten Arbeitsschritt, also ohne Volltextprüfung, entsprechend der vordefinierten Einschlusskriterien als eindeutig irrelevant ausgeschlossen werden, weil bereits aus den jeweiligen Abstracts zweifelsfrei hervorging, dass es sich bei der entsprechenden Studie nicht um eine RCT handelte. Die 108 in diesem ersten Arbeitsschritt nicht aussortierten

Publikationen wurden für den nächsten Schritt des Selektionsprozesses im Volltext beschafft.

Darüber hinaus gingen die 72 über die Vorgängerdissertation identifizierten Artikel, die 42 nachträglich identifizierten Publikationen (engl. reference tracking) sowie der über die Publikationslisten gefundene Artikel in die Volltextprüfung ein, sodass insgesamt 223 Volltexte durch den Autor der vorliegenden Arbeit formal durchgesehen wurden.

Im Rahmen der Volltextprüfung konnten weitere 128 Publikationen ausgeschlossen werden, sodass 95 Artikel zurückblieben, die sich auf 60 RCTs bezogen (hierzu zählten sowohl noch nicht abgeschlossene RCTs, die sich im Stadium der Planung oder Durchführung befanden und für die folglich noch keine Ergebnisse vorlagen, als auch bereits abgeschlossene RCTs mit vollständigen Ergebnissen). Die Differenz zwischen der Anzahl an Studien und der Anzahl an Publikationen erklärt sich dadurch, dass es nicht selten mehrere Publikationen zu einer Studie gab.

Die o. g. 95 Arbeiten setzten sich folgendermaßen zusammen:

- Hauptpublikationen: 60 (davon 17 eigenständige Studienprotokolle)
- Zusatzpublikationen: 35

In der vorliegenden Arbeit gab es demnach im Durchschnitt 1,5 Publikationen pro Studie mit einer Spannweite von 1 bis 5 Artikeln pro Studie.

Zusammenfassend kann somit gesagt werden, dass am Ende des Selektionsprozesses 60 RCTs, beschrieben in 95 relevanten Publikationen, die ursprünglich festgelegten Einschlusskriterien erfüllten (Abbildung 4).

#### **3.1.4 Nachträglicher Studienausschluss**

Nach genauer Prüfung der 60 Studien (in 95 Publikationen) erfolgte ein nachträglicher Ausschluss von 10 Studien (in 16 Publikationen), sodass 50 Studien (in 79 Publikationen) in die nachfolgende engere Analyse (3.2) eingingen.<sup>1</sup>

Dieser nachträgliche Studienausschluss wurde durch den Autor der vorliegenden Arbeit und dessen Betreuer als sinnvoll erachtet, da durch die ursprünglich definierten Einschlusskriterien auch Studien eingeschlossen worden waren, bei denen es sich eindeutig *nicht* um allgemeinmedizinische Forschung bzw. Forschung von *deutschen* allgemeinmedizinischen Instituten handelte.

---

<sup>1</sup> Die genaue Zuordnung von Publikationen zu Studien kann der „Übersicht aller Studienreferenzen“ im Literaturverzeichnis dieser Arbeit entnommen werden.

Dennoch sollte ein solcher nachträglicher Ausschluss von Studien bzw. Publikationen, die eigentlich die Einschlusskriterien erfüllten, nicht leichtfertig erfolgen. Damit klar nachvollziehbar bleibt, welche Studien nachträglich aus der Analyse genommen wurden und was die Gründe für deren Ausschluss waren, findet sich in Tabelle 10 eine detaillierte Auflistung.

In Tabelle 10 sind die nachträglich ausgeschlossenen Studien nach den folgenden drei möglichen Ausschlussgründen gruppiert:

- a) Erst- und/oder Letztautoren hatten eine Affiliation zur deutschen universitären Allgemeinmedizin zum Zeitpunkt der Publikation ( $n = 4$ ). Planung und Durchführung der entsprechenden Studie fanden allerdings in einem anderen, „Allgemeinmedizin-fremden“ Fachbereich statt.
- b) Die RCTs waren Kurzzeitstudien, die physiologische Messgrößen als primäre Zielparameter aufwiesen und somit der experimentellen Forschung zuzurechnen sind ( $n = 5$ ). Es fehlte der Bezug zur Allgemeinmedizin im engeren Sinne.
- c) Der Letztautor hatte eine Affiliation zu einem deutschen Institut für Allgemeinmedizin, die Studie wurde aber im Ausland und unter Federführung einer ausländischen Institution durchgeführt ( $n = 1$ ). Somit fehlte der Bezug zur *deutschen* universitären Allgemeinmedizin.

Drei Publikationen der Gruppe b) erfüllten zusätzlich den Ausschlussgrund a) [Gavrylyuk et al., 2010, Meissner et al., 2011, Ronel et al., 2011].

Tabelle 10: Ausgeschlossene Studien (n = 10)

Für Zuordnung von Studien zu Publikationen siehe „Übersicht aller Studienreferenzen“ im Literaturverzeichnis.

Erstautor Jahr	Teilnehmer/Erkrankung	Intervention(en)	Vergleichsgruppe	Zentrale Zielkriterien	Random.	Fallzahl*
<b>Erstautoren mit Anbindung an Allgemeinmedizininstitut zum Zeitpunkt der Publikation, aber Studienplanung und -durchführung fand in anderer Einrichtung statt</b>						
Joos 2005	Morbus Crohn	Akupunktur	Minimal-Akupunktur**	Krankheitsaktivität (CDAI)	Pat./Prob.	51
Joos 2006	Colitis ulcerosa	Akupunktur + Moxibustion	Sham-Akupunktur**	Krankheitsaktivität (CAI)	Pat./Prob.	29
Schneider 2005	Postoperative Übelkeit und Erbrechen (PONV)	Akupunktur	Sham-Akupunktur**	Übelkeit, Erbrechen, subj. Körperwahrnehmung	Pat./Prob.	220
Schneider 2006	Reizdarmsyndrom	Akupunktur	Sham-Akupunktur**	Lebensqualität (QoL)	Pat./Prob.	43
<b>experimentelle Kurzzeitstudien mit physiologischen Messgrößen als Zielparameter</b>						
Doering 2001	Gesunde Probanden	Gesichtsguss + Nackenwickel (kalt)	Gesichtsguss + Nackenwickel (warm)	Kognitive Hirnfunktion	Pat./Prob. (c-o)	24
Doering 2002	Gesunde Probanden	CO <sub>2</sub> -Nassapplikation	CO <sub>2</sub> -Trockenapplikation	Flussgeschwindigkeit in A. cerebri media	Pat./Prob.	22
Gavrylyuk 2010	Gesunde Probanden	Verbal induzierte Erwartungen bzgl. Pupillenweite + Placebo/Verum-Augentropfen (4 Möglichkeiten)	Ankündigung „kein Effekt auf Pupillenweite“ + Placeboaugentropfen	Pupillenweite und Akkomodation	Pat./Prob.	50
Meissner 2011	Gesunde Probanden	I1: Globuli + verbale Ankündigung „Blutdruck sinkt“ I2: Placebo + verbale Ankündigung „Blutdruck sinkt“	Keine Intervention	Blutdruck, Herzfrequenz, Hautleitfähigkeit	Pat./Prob.	45
Ronel 2011	Koronarangiographiepatienten	Placebo + verbale Suggestion „Herzkranzgefäße weiten sich“	Placebo ohne verbale Suggestion	Durchmesser der Herzkranzarterien	Pat./Prob.	30
<b>Letztautor mit Anbindung zu einem deutschen Institut für Allgemeinmedizin, aber Studie wurde im Ausland durchgeführt</b>						
Ludman 2013	Depression + Diabetes oder Depression + KHK	Multimodales Interventionsprogramm durch MFAs	Standardtherapie	Depressive Symptome, HbA1c, Lebensqualität (QoL), Blutdruck	Pat./Prob.	214

\* die Fallzahl ist die Anzahl der Patienten/Probanden, die an der entsprechenden Studie teilgenommen haben und in die statistische Auswertung der Studie miteingegangen sind

\*\* Minimal- oder Sham-Akupunktur dient bei Akupunktur-Studien häufig als Kontrollintervention. Dabei erfolgt im Vergleich zur echten Akupunktur z. B. keine (Placebo-Nadeln) bzw. nur eine oberflächliche Hautpenetration oder es werden „falsche“ Akupunkturpunkte bzw. Nicht-Akupunkturpunkte verwendet. Die verwendeten Methoden sind allerdings sehr heterogen.

Abkürzungen (weitere Details zu den Abkürzungen finden sich im Abkürzungsverzeichnis):

A. = Arteria; bzgl. = bezüglich; CAI = colitis activity index; CDAI = crohn's disease activity index; c-o = cross-over Studie; CO<sub>2</sub> = Kohlenstoffdioxid; Diabetes = Diabetes mellitus; HbA1c = glykisiertes Hämoglobin; 1/1/2 = Interventionsgruppe 1/2; KHK = koronare Herzkrankheit; MFA = medizinische Fachangestellte; Pat. = Patient(en); PONV = postoperative nausea and vomiting; Prob. = Proband(en); Random. = Ebene, auf der die Randomisierung stattgefunden hat; subj. = subjektiv(e)

## **3.2 Merkmale der eingeschlossenen Studien**

50 RCTs (in 79 Publikationen) wurden in die Analyse eingeschlossen (Abbildung 4). Diese setzten sich folgendermaßen zusammen:

- 33 Studien (in 62 Publikationen) mit (zumindest ersten) Ergebnissen
- 17 Studienprotokolle (in 17 Publikationen), zu denen bis Ende 2014 noch keine Ergebnisse publiziert worden waren

Auf diese 50 Studien (in 79 Publikationen) bezieht sich die nachfolgende Auswertung.

### **3.2.1 Zeitpunkt der Veröffentlichung, Publikationssprache und Autorenschaft**

Wie unter 2.5.1 beschrieben, stellte jeweils eine RCT (unabhängig von der Anzahl der ihr zugehörigen Publikationen) die Analyseeinheit für die vorliegende Arbeit dar. Allerdings war die Erhebung einiger bibliographischer Merkmale auf der Ebene der 50 eingeschlossenen RCTs nicht möglich, sodass hier – im Gegensatz zur Analyse der weiteren Merkmale 3.2.2 bis 3.2.7 – auf die Ebene der 79 Publikationen ausgewichen wurde.

In Tabelle 11 und in Abbildung 5 sind die 79 eingeschlossenen Artikel nach dem Jahr ihrer Veröffentlichung eingeteilt. Hierbei fällt auf, dass von 2000 bis 2002 kein Artikel, von 2012 bis 2014 jedoch 35 Artikel publiziert wurden. Es zeigt sich also von der Jahrtausendwende bis heute eine deutliche Zunahme der Veröffentlichungen zu mit der deutschen universitären Allgemeinmedizin in Verbindung stehenden RCTs. Besonders eindrücklich zeigt sich dieser Trend, wenn man nur die Studienprotokolle betrachtet: Sie stiegen von einem Artikel im Zeitraum 2006 bis 2008 auf 14 Artikel im Zeitraum 2012 bis 2014 an.

69 (87 %) der 79 Artikel waren in englischer, 10 (13 %) in deutscher Sprache verfasst.

Das Einschlusskriterium 2.2.3 („Erst- und/oder Letztautor hat Affiliation zur deutschen universitären Allgemeinmedizin“) wurde von 74 (94 %) der 79 eingeschlossenen Artikel erfüllt, 5 Publikationen (6 %) erfüllten es nicht; sie wurden dennoch aufgenommen, weil sie wichtige Zusatzinformationen zu RCTs enthielten, die bereits zuvor eingeschlossen worden waren. Bei 70 (89 %) der 79 eingeschlossenen Artikel hatte der Erstautor, bei 57 (72 %) hatten sogar Erst- und Letztautor eine Affiliation zur deutschen universitären Allgemeinmedizin. Dies legt nahe, dass der Großteil der eingeschlossenen RCTs federführend an allgemeinmedizinischen Instituten geplant und durchgeführt wurde. Studien, die primär von anderen Abteilungen organisiert wurden, bei denen also Allgemeinmediziner „nur“ am Rande mitgewirkt haben, können sich zwar unter den eingeschlossenen RCTs befinden, stellen aber höchstwahrscheinlich eine Minderheit dar.

Eine weitere Auffälligkeit in Bezug auf die Autorenschaft stellt sowohl die (median) höhere Anzahl der Gesamtautoren (9 vs. 7) als auch der allgemeinmedizinischen Autoren (6 vs. 4) bei den Studienprotokollen gegenüber den Publikationen mit Ergebnissen dar.

Ob Autoren von Publikationen zu eingeschlossenen RCTs eine internationale Affiliation oder eine Affiliation zu einem Institut für medizinische Statistik oder Biometrie hatten, wurde auf Ebene der 50 Studien und nicht – wie bei der allgemeinmedizinischen Autorenschaft – für die 79 Publikationen überprüft (Tabelle 12). Denn selbst wenn ein entsprechender Autor nur an *einer* von mehreren Publikationen zu einer RCT beteiligt war, konnte davon ausgegangen werden, dass dieser Autor an der Studie mitgewirkt hat.

Es zeigte sich bei 31 (62 %) der 50 eingeschlossenen Studien eine Mitwirkung durch ein Institut für Statistik/Biometrie. Bei den Studien mit Ergebnissen waren in 52 % der Fälle (17 von 33), bei den Studienprotokollen in 82 % der Fälle (14 von 17) Biometriker beteiligt.

An 15 (30 %) der 50 Studien waren Autoren aus dem Ausland beteiligt. Unter den 33 Studien, zu denen bereits Ergebnisse publiziert worden waren, fanden sich 11 (33%) mit internationaler Beteiligung. Bei den sich in Planung oder Durchführung befindlichen RCTs ohne bereits veröffentlichte Ergebnisse (Studienprotokolle) gab es hingegen nur in 4 (24 %) von 17 Fällen (einen) internationale(n) Autor(en).

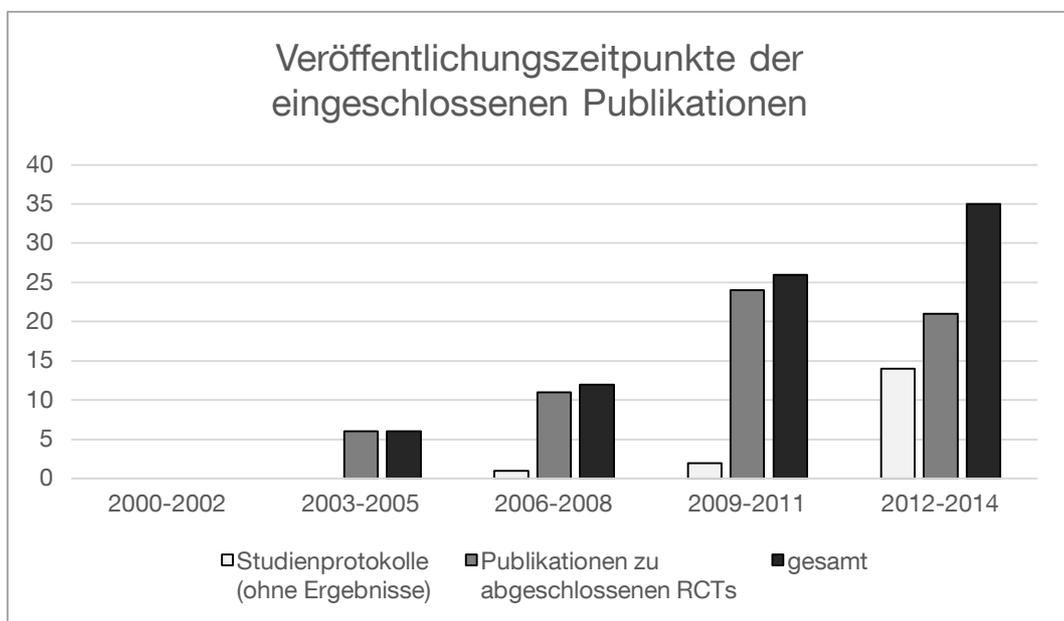


Abbildung 5: Gruppierung der 79 eingeschlossenen Publikationen nach dem Zeitpunkt ihrer Veröffentlichung (in 3-Jahres-Abschnitten)

Gezeigt ist eine Gesamtbetrachtung sowie eine getrennte Darstellung der Publikationen zu Studienprotokollen (ohne Ergebnisse) und der Publikationen zu bereits abgeschlossenen RCTs.

### 3 Ergebnisse

Tabelle 11: Veröffentlichungszeiträume, Publikationssprache und Autoren

	Publikationen zu RCTs mit Ergebnissen n = 62	Publikationen zu RCTs ohne Ergebnisse (Studienprotokolle) n = 17	Gesamt n = 79
<b>Veröffentlichungszeiträume</b>			
- 2000 bis 2002	-	-	- (0 %)
- 2003 bis 2005	6	-	6 (8 %)
- 2006 bis 2008	11	1	12 (15 %)
- 2009 bis 2011	24	2	26 (33 %)
- 2012 bis 2014	21	14	35 (44 %)
<b>Publikationssprache</b>			
- Deutsch	10	-	10 (13 %)
- Englisch	52	17	69 (87 %)
<b>Autoren</b>			
- Gesamtzahl Autoren pro Publikation	7 (2–13)	9 (5–17)	7 (2–17)
- Zahl Autoren mit Affiliation zur AM	4 (0–9)	6 (1–14)	4 (0–14)
- Erstautor aus AM	53	17	70 (89 %)
- Letztautor aus AM	47	14	61 (77 %)
- Erst- und Letztautor aus AM	43	14	57 (72 %)
- weder Erst- noch Letztautor AM	5	-	5 (6 %)

Abkürzungen (weitere Details zu den Abkürzungen finden sich im Abkürzungsverzeichnis):

AM = Allgemeinmedizin

Tabelle 12: Internationale Autoren oder Mitautoren aus der Biometrie

	RCTs mit Ergebnissen n = 33	RCTs ohne Ergebnisse (Studienprotokolle) n = 17	Gesamt n = 50
Mitautor aus Biometrie	17	14	31 (62 %)
Mitautor international	11	4	15 (30 %)

### 3.2.2 Beteiligte Abteilungen

Eine deutsche Universitätsabteilung für Allgemeinmedizin wurde als „an einer Studie beteiligt“ betrachtet, wenn bei mindestens einer Publikation zu dieser Studie ein Autor (unabhängig davon, ob dieser Erst-, Mittel- oder Letztautor war) eine Affiliation zu der entsprechenden allgemeinmedizinischen Abteilung vorweisen konnte.

Es zeigte sich, dass 34 (68 %) der 50 eingeschlossenen Studien von nur *einem* Institut betreut wurden (Tabelle 13). Bei 16 (32 %) der 50 eingeschlossenen Studien waren zwei oder mehr universitäre Allgemeinmedizinabteilungen aus Deutschland beteiligt.

An den 50 in der vorliegenden Arbeit untersuchten randomisierten Studien waren insgesamt 15 (42 %) der heute 36 allgemeinmedizinischen Universitätsabteilungen in Deutschland beteiligt.

In Tabelle 13 werden die 15 allgemeinmedizinischen Universitätsabteilungen in Deutschland, die an den 50 eingeschlossenen Studien beteiligt waren, namentlich genannt. Da mehrere Institute an einer Studie beteiligt sein konnten, sind an dieser Stelle Mehrfachnennungen möglich.

Bei 39 (78 %) RCTs, also bei über drei Viertel der 50 eingeschlossenen Studien, war jeweils mindestens eine der sechs allgemeinmedizinischen Abteilungen der Universitäten Heidelberg, Frankfurt, Witten-Herdecke, Göttingen, München (TUM) und Düsseldorf beteiligt. Demnach ist eine verhältnismäßig kleine Gruppe von allgemeinmedizinischen Universitätsabteilungen in Deutschland an einem Großteil der durchgeführten bzw. geplanten allgemeinmedizinischen RCTs beteiligt, während viele andere allgemeinmedizinische Universitätsabteilungen bisher an keiner RCT beteiligt waren.

Unter den 33 eingeschlossenen Studien mit Ergebnissen befinden sich 27 (82 %), bei denen jeweils mindestens eines der fünf allgemeinmedizinischen Institute der Universitäten Heidelberg, Frankfurt, Witten-Herdecke, München (TUM) und Göttingen beteiligt war.

Betrachtet man die 17 Studienprotokolle (ohne Ergebnisse), so zeigt sich, dass bei 14 (82 %) mindestens eine der fünf Allgemeinmedizininstitute der Universitäten Rostock, Düsseldorf, Hamburg, Heidelberg und Jena beteiligt war.

### 3 Ergebnisse

Tabelle 13: Beteiligung der einzelnen allgemeinmedizinischen Institute an den eingeschlossenen RCTs

	RCTs mit Ergebnissen n = 33	RCTs ohne Ergebnisse (Studienprotokolle) n = 17	Gesamt n = 50
<b>Anzahl beteiligter Abteilungen</b>			
- eine	23	11	34 (68 %)
- zwei	6	3	9 (18 %)
- drei	3	3	6 (12 %)
- vier	1	-	1 (2 %)
<b>beteiligte allgemeinmedizinische Institute deutscher Universitäten*</b>			
- Heidelberg	8	3	11 (22 %)
- Frankfurt	8	2	10 (20 %)
- Witten	6	1	7 (14 %)
- Göttingen	5	2	7 (14 %)
- München (TUM)	6	-	6 (12 %)
- Düsseldorf	2	4	6 (12 %)
- Hamburg	2	3	5 (10 %)
- Jena	2	3	5 (10 %)
- Rostock	-	5	5 (10 %)
- Hannover	2	2	4 (8 %)
- Leipzig	2	-	2 (4 %)
- Marburg	2	-	2 (4 %)
- Freiburg	1	1	2 (4 %)
- Dresden	1	-	1 (2 %)
- Greifswald	1	-	1 (2 %)
- gesamt	48	26	74 (148 %)

\* An dieser Stelle sind Mehrfachnennungen möglich.

Abkürzungen (weitere Details zu den Abkürzungen finden sich im Abkürzungsverzeichnis):

TUM = Technische Universität München (in Abgrenzung zur Ludwig-Maximilians-Universität München)

#### 3.2.3 Finanzierung

Bei 8 (16 %) der 50 eingeschlossenen Studien fanden sich keine Informationen zu deren Finanzierung. Für die 42 (84 %) verbleibenden Studien erfolgte eine Kategorisierung entsprechend der Herkunft der finanziellen Mittel (die 8 Kategorien sind in Tabelle 14 genannt). Von diesen 42 Studien bezogen 36 (72 %) finanzielle Mittel aus nur einer, 6 (12 %) finanzielle Mittel aus mehr als einer der 8 Kategorien.

25 (50 %) der 50 eingeschlossenen Studien erhielten finanzielle Mittel durch das Bundesministerium für Bildung und Forschung (BMBF), bei 21 (42 %) Studien war das BMBF sogar der alleinige Geldgeber (die 4 verbleibenden Studien erhielten zusätzlich finanzielle Zuwendungen aus anderen Quellen). 4 (8 %) der eingeschlossenen Studien erhielten finanzielle Zuwendungen durch das Bundesgesundheitsministerium (BMG), in 3 (6 %) dieser Fälle gab es neben dem BMG keine weiteren Geldgeber. Somit war bei 24 (48 %) Studien ein Bundesministerium der alleinige Geldgeber.

In 7 (14 %) Studien erfolgte die Finanzierung aus den Eigenmitteln des durchführenden Instituts für Allgemeinmedizin oder der zugehörigen Universität. 4 dieser Studien erhielten allerdings zusätzliche Gelder aus anderen Quellen.

5 (10 %) der eingeschlossenen Studien wurden von Krankenkassen finanziert oder teilfinanziert. In 4 (8 %) Fällen handelte es sich dabei um die AOK (Bundesverband oder Kombination aus Bundes- und Landesverband). 2 (4 %) Studien wurden sogar ausschließlich durch die AOK finanziert.

Bei den drei unter „sonstige“ vermerkten Geldgebern handelte es sich um die Europäische Union [Jäger et al., 2013], das Bayerische Staatsministerium für Umwelt und Gesundheit [Freiberger et al., 2013] sowie um einen Zusammenschluss aus dem Bundesministerium für Arbeit und Soziales mit der deutschen Rentenversicherung [Schmidt et al., 2010]. Dieser Zusammenschluss bildete gemeinsam mit dem BMBF, dem BMG sowie verschiedenen Krankenkassen den Förderschwerpunkt „Versorgungsnahe Forschung – Chronische Krankheiten und Patientenorientierung“, welcher eine der eingeschlossenen Studien finanzierte [Schmidt et al., 2010].

Keine der untersuchten Publikationen enthielt Informationen über konkrete Beträge der finanziellen Zuwendungen. In aller Regel war es bei mehreren Geldgebern auch nicht möglich, die relativen Anteile, die durch die verschiedenen Geldgeber beigesteuert wurden, nachzuvollziehen. Rückschlüsse auf die Höhe der Fördergelder bzw. auf das Verhältnis der Fördergelder der verschiedenen Quellen zueinander, sind deshalb nicht möglich.

### 3 Ergebnisse

Tabelle 14: Finanzierung der eingeschlossenen Studien

Finanzmittelgeber	RCTs mit Ergebnissen n = 33	RCTs ohne Ergebnisse (Studienprotokolle) n = 17	Gesamt n = 50
BMBF	12	13	25 (50 %)
BMG	2	2	4 (8 %)
Eigenmittel Institut/Universität	5	2	7 (14 %)
Krankenkassen	3	2	5 (10 %)
Industrie	3	-	3 (6 %)
Stiftungen/Gesellschaften/Vereine	2	1	3 (6 %)
DFG	1	1	2 (4 %)
sonstiges	1	2	3 (6 %)
unklar	8	-	8 (16 %)
gesamt	37	23	60 (120 %)

In dieser Tabelle sind Mehrfachnennungen möglich.

Abkürzungen (weitere Details zu den Abkürzungen finden sich im Abkürzungsverzeichnis):

BMBF = Bundesministerium für Bildung und Forschung; BMG = Bundesministerium für Gesundheit; DFG = Deutsche Forschungsgemeinschaft

#### 3.2.4 Untersuchte Fragestellungen

Die Diversität der Fragestellungen der in der vorliegenden Arbeit untersuchten randomisierten Studien war groß. Es konnte deshalb nur eine grobe Einteilung in Bezug auf die Art der jeweils untersuchten Intervention (3.2.4.1) und hinsichtlich eines möglichen Krankheitsbezugs (0) vorgenommen werden. Studien, die einen direkten Krankheitsbezug vorwiesen, wurden zudem anhand gängiger Klassifikationssysteme kategorisiert (3.2.4.3). Zunächst soll jedoch anhand der in Beispiel 7 vorgestellten eingeschlossenen RCTs, ein Überblick über die Spannweite der untersuchten Fragestellungen gegeben werden.

Beispiel 7:

- A Altiner et al. [2012b] (Studienprotokoll):  
*Führt ein Ärztetraining zum Thema patientenzentrierte Kommunikation bei älteren Patienten mit  $\geq 3$  chronischen Erkrankungen im Vergleich zu usual care (keine Intervention) zu Unterschieden in der Medikation und einer besseren Lebensqualität? (clusterrandomisiert, 600 Patienten angestrebt)*

- B Bozorgmehr et al. [2014] (Studienprotokoll):  
*Führt eine „Case-management“-Intervention im Vergleich zu usual care (keine Intervention) bei Patienten mit Diabetes mellitus Typ 2 und  $\geq 2$  weiteren schweren chronischen Erkrankungen zu einer Verbesserung von Selbstversorgungsverhaltensweisen und anderen Krankheitsindikatoren? (clusterrandomisiert, 582 Patienten angestrebt)*
- C Gágyor et al. [2012] (Studienprotokoll):  
*Reduziert eine initiale Therapie mit Ibuprofen bei Frauen mit unkompliziertem Harnwegsinfekt im Vergleich zu einer sofortigen Antibiotikatherapie die Zahl der Antibiotikaverschreibungen in einem 28-Tage-Zeitraum, ohne dass es zu einer stärkeren Krankheitsbelastung kommt? (494 Patienten angestrebt)*
- D Hoffmann et al. [2014]:  
*Führt eine definierte Intervention mit Praxisteam-Schulungen im Vergleich zu einer Minimal-Intervention zu einer Verbesserung des Fehlermanagements und einer Reduktion von kritischen Vorfällen? (65 Praxisteams)*
- E Jobst et al. [2005]:  
*Unterscheidet sich die Zahl der Krankheitstage in einem Zeitraum von 20 Wochen bei Patienten mit chronisch rezidivierenden Infekten der Atemwege, die entweder intramuskuläre Eigenblutgaben oder ein homöopathisches Komplexmittel erhalten? (80 Patienten)*
- F Mehring et al. [2013]:  
*Kommt es bei übergewichtigen Patienten ( $BMI \geq 25$ ) unter einer web-basierten, verhaltenstherapeutischen Intervention zu einer größeren Gewichtsreduktion als bei Patienten, denen eine solche Intervention nicht angeboten wird? (clusterrandomisiert, 186 Patienten)*

### 3.2.4.1 Interventionsart

Unter den eingeschlossenen randomisierten Studien kristallisierten sich zwei große Arten von untersuchten Interventionen heraus. Dies waren zum einen Interventionen, die spezifische diagnostische oder therapeutische Maßnahmen (z. B. ein Medikament, eine physiotherapeutische Übung oder ein naturheilkundliches Behandlungsverfahren; vgl. Studien C und E in Beispiel 7) untersuchten (diese Art der Intervention wird im weiteren Verlauf als „Intervention zu einer definierten Behandlung“ bezeichnet), zum anderen Interventionen, die über Veränderungen von Abläufen (z. B. über Edukations-, Trainings- oder „Case-management“-Programme; vgl. Studien A, B, D und F in Beispiel 7) auf die Optimierung von Vorgehensweisen in der Praxis oder durch den Patienten abzielten (diese Interventionsart wird im Folgenden „prozessorientierte“ Intervention genannt). War es nicht möglich, die in einer eingeschlossenen RCT angewendete Intervention einer der beiden genannten Interventionsarten zuzuordnen, so erfolgte die Zuteilung in die Kategorie „andere“.

Insgesamt befassten sich 31 (62 %) der 50 eingeschlossenen Studien mit „prozessorientierten“ Interventionen, während 14 (28 %) RCTs Interventionen zu „einer definierten Behandlung“ untersuchten (Tabelle 15). Unter den 33 RCTs mit Ergebnissen befanden sich 16 (48 %) Studien, die „prozessorientierte“ Interventionen untersuchten. Mit 15 (88 %) Studien zu „prozessorientierten“ Interventionen war dieser Anteil bei den 17 Studienprotokollen (ohne Ergebnisse) deutlich höher.

Tabelle 15: Einteilung der eingeschlossenen Studien nach Art der Intervention

Interventionsart	RCTs mit Ergebnissen n = 33	RCTs ohne Ergebnisse (Studienprotokolle) n = 17	Gesamt n = 50
„definierte Behandlung“	12	2	14 (28 %)
„prozessorientierte“ Intervention	16	15	31 (62 %)
andere	5	-	5 (10 %)

### 3.2.4.2 Krankheitsbezug

Bei 34 (68 %) der 50 eingeschlossenen Studien bezog sich die Fragestellung auf eine einzelne Erkrankung (z. B. Rückenschmerzen, Asthma bronchiale, Diabetes mellitus, Depression, Harnwegsinfekt, Hypertonie o. ä.; vgl. Studien B, C, E und F Beispiel 7), auf eng miteinander verwandte Diagnosen (z. B. ähnliche psychiatrische Diagnosen, wie

„Patienten mit Angsterkrankung, Depression oder somatoformer Störung“) oder eng umschriebene andere Indikationen (z. B. Studien zum Thema Raucherentwöhnung oder Mammographie). Dies wird im Folgenden als „direkter Krankheitsbezug“ bezeichnet.

Bei 9 (18 %) Studien lag ein Krankheitsbezug vor, jedoch war keine Eingrenzung auf eine einzelne Diagnose möglich; vielmehr handelte es sich um Krankheits- bzw. Symptomgruppen. Hierzu zählten z. B. Studien, bei denen multimorbide geriatrische Patienten untersucht werden sollten und somit Alter und eine bestimmte Anzahl an diagnostizierten chronischen Erkrankungen als Einschlusskriterien galten (vgl. Studie A in Beispiel 7).

7 (14 %) Studien wiesen keinen oder keinen unmittelbaren Krankheitsbezug auf, wie dies auch in Studie D (Beispiel 7) der Fall ist. In erster Linie waren dies Studien, die den Wissenszuwachs bei Ärzten oder Medizinstudenten durch verschiedene Interventionen untersuchten. Tabelle 16 gibt einen Überblick über den Krankheitsbezug der eingeschlossenen Studien.

Tabelle 16: Einteilung der eingeschlossenen Studien nach Krankheitsbezug

Krankheitsbezug	RCTs mit Ergebnissen n = 33	RCTs ohne Ergebnisse (Studienprotokolle) n = 17	Gesamt n = 50
direkter Krankheitsbezug	23	11	34 (68 %)
Krankheits-/Symptomgruppe	3	6	9 (18 %)
keiner bzw. kein unmittelbarer	7	-	7 (14 %)

### 3.2.4.3 Einteilung der untersuchten Krankheiten

Diejenigen Studien, welche einen „direkten Krankheitsbezug“ vorwiesen (0), wurden anhand der internationalen Klassifikationssysteme ICPC-2 [ICPC-2, 2011] und ICD-10 [ICD-10, 2016] kategorisiert (Tabelle 17). Die vier am häufigsten genannten ICPC-2-Überkategorien waren „Psychologisch“ (7 Studien, 21 %), „Kreislauf“ (6 Studien, 18 %), „Bewegungsapparat“ (6 Studien, 18 %) und „Atmungsorgane“ (6 Studien, 18 %). Dieselben bzw. die entsprechenden Kapitel waren auch in der ICD-10-Klassifikation die meistgenannten.

### 3 Ergebnisse

Tabelle 17: Einteilung der Studien mit direktem Krankheitsbezug (n = 34) nach ICPC-2 und ICD-10

Untersuchte Krankheitskategorien (nach ICPC-2/ICD-10)		RCTs mit Ergebnissen n = 23	RCTs ohne Ergebnisse (Studienprotokolle) n = 11	Gesamt n = 34
<b>Einteilung nach ICPC-2</b>				
P	Psychologisch	5	2	7 (21 %)
K	Kreislauf	3	3	6 (18 %)
L	Bewegungsapparat	5	1	6 (18 %)
R	Atmungsorgane	5	1	6 (18 %)
T	Endokrin, metabolisch, Ernährung	2	2	4 (12 %)
U	Urologisch	2	1	3 (9 %)
N	Neurologisch	1	-	1 (3 %)
X	Weibliches Genitale/Brust	-	1	1 (3 %)
<b>Einteilung nach ICD-10</b>				
F00-F99	Psychische und Verhaltensstörungen	5	2	7 (21 %)
M00-M99	Muskuloskeletale und Bindegewebserkrankungen	5	1	6 (18 %)
I00-I99	Krankheiten des Kreislaufsystems	3	2	5 (15 %)
J00-J99	Krankheiten des Atmungssystems	4	1	5 (15 %)
E00-E90	Endokrine, Ernährungs- und Stoffwechselkrankheiten	2	2	4 (12 %)
N00-N99	Krankheiten des Urogenitalsystems	2	1	3 (9 %)
G00-G99	Krankheiten des Nervensystems	1	-	1 (3 %)
R00-R99	Anderenorts nicht klassifizierte Symptome und Befunde	1	-	1 (3 %)
A00-B99	Bestimmte infektiöse und parasitäre Erkrankungen	-	1	1 (3 %)
C00-D48	Neubildungen	-	1	1 (3 %)

Da die ICPC-2- und ICD-10-Klassifikation nur eine grobe Zuordnung der eingeschlossenen Studien zu übergeordneten Kapiteln erlauben, werden in Tabelle 18 alle Studienthemen, die mindestens zweimal untersucht wurden, im Freitext aufgelistet.

Tabelle 18: Untersuchte Krankheiten im Freitext

Untersuchte Krankheiten	RCTs mit Ergebnissen n = 23	RCTs ohne Ergebnisse (Studienprotokolle) n = 11	Gesamt n = 34
Atemwegsinfekt/Husten/Erkältung	3	1	4 (12 %)
Depression/Angst-/somatoforme Störung	2	2	4 (12 %)
Rückenschmerz/Nackenschmerz	3	1	4 (12 %)
Diabetes mellitus Typ II	2	1	3 (9 %)
Arthrose	2	-	2 (6 %)
Asthma bronchiale/COPD	2	-	2 (6 %)
Demenz	2	-	2 (6 %)
Harnwegsinfekt	1	1	2 (6 %)
Herzinsuffizienz	2	-	2 (6 %)
Hypertonie	1	1	2 (6 %)
sonstiges	3	4	7 (21 %)

Abkürzungen (weitere Details zu den Abkürzungen finden sich im Abkürzungsverzeichnis):  
COPD = Chronic Obstructive Pulmonary Disease

#### 3.2.4.4 „PICO-Tabellen“

Die in den Tabellen 19 bis 21 gezeigte Zusammenschau der 50 eingeschlossenen RCTs ermöglicht einen Überblick über die Patienten/Probanden, die untersuchten Interventionen/Kontrollen und die zentrale Zielkriterien (PICO-Fragestellung, [PICO, 2016]) einer jeden eingeschlossenen Studie. Des Weiteren werden Informationen zur Randomisierung sowie Fallzahlen genannt. Wurden in einer Studie mehrere Fragestellungen behandelt, so wurde die jeweils im Vordergrund stehende, primäre Fragestellung vermerkt. Es wurde dabei folgende Einteilung vorgenommen:

In Tabelle 19 und 20 sind die 33 abgeschlossenen RCTs aufgelistet. Sie sind unterteilt in die 17 abgeschlossenen clusterrandomisierten RCTs (Tabelle 19) und die 16 abgeschlossenen „klassischen“ RCTs (Tabelle 20).

Die 17 RCTs, zu denen nur Studienprotokolle ohne Ergebnisse vorlagen (die sich also noch im Planungs- oder Durchführungsstadium befanden), sind in Tabelle 21 aufgelistet.

Tabelle 19: Abgeschlossene RCTs – clusterrandomisiert (n = 17)

Für Zuordnung von Studien zu Publikationen siehe „Übersicht aller Studienreferenzen“ im Literaturverzeichnis.

Erstautor Jahr	Teilnehmer/Erkrankung	Intervention(en)	Vergleichsgruppe	Zentrale Zielkriterien	Random.	Fallzahl*
Altiner 2007	Husten	Ärzte- und Patientenschulung zu Antibiotika	Keine Intervention	Anzahl verschriebener Antibiotika	Arzt/Praxis	104/2787
Becker 2008	Rückenschmerz	11: Ärzteschulung zu leitliniengerechter Kreuzschmerztherapie 12: I1 + Training der MFAs zu Pat.-Motivation	Leitlinie Kreuzschmerz per Post an Praxen	durch Rückenschmerzen verursachte Funktionsbeeinträchtigungen (gemessen mittels FFbH-R)	Arzt/Praxis	118/1378
Erlor 2012	Nierensuffizienz	Ärzte- und Pat.-Schulung zu Medikamentendosierung bei Nierensuffizienz	Standardtherapie	Anzahl an zu hohen Dosierungen	Arzt/Praxis	46/404
Freiberger 2013	Sturzgefährdung	Trainingsprogramm zur Sturzprävention	Keine Intervention	Sturzrisiko**	Arzt/Praxis	33/378
Gensichen 2009	Depression	Schulung der MFAs, danach „Case-management“ durch MFAs	Keine Intervention	Depressive Symptomatik	Arzt/Praxis	74/626
Junius-Walker 2012	Hausarztpatienten	Ärzteschulung + Leitfaden PrefCheck	Keine Intervention	Grad der Arzt-Patienten-Übereinkunft bzgl. Priorisierung von Gesundheitsproblemen	Arzt/Praxis	42/347
Kaufmann-Kolle 2011	Asthma bronchiale	Ärztliche QZ mit Benchmark (d. h. QZ-Ärzte vergleichen die Werte ihrer Asthmapatienten)	Ärztliche QZ ohne Benchmark	Ungünstige vs. leitliniengerechte Medikation	QZ/CME	6/unklar
Krones 2008	Hausarztpat. (nach Cholesterinmessung)	Ärzteschulung zu shared-decision-making (SDM) bei KHK-Risikopatienten	Ärzteschulung zu anderen Themen	Patientenbeteiligung und -zufriedenheit	QZ/CME	14/1132
Mehring 2013	Patienten mit BMI $\geq$ 25	Online-Abnehmprogramm	Standardtherapie	Gewichtsreduktion nach 12 Wo.	Arzt/Praxis	92/186
Mehring 2014	Pat. mit Wunsch nach Rauchentwöhnung	Internetbasiertes Programm zur Rauchentwöhnung	Standardtherapie	Biochemisch nachgewiesener Raucherstatus nach 12 Wo.	Arzt/Praxis	92/168
Peters-Klimm 2009	Herzinsuffizienz	Ärzteschulung zum Thema Herzinsuffizienz	Kurzvorlesung für Ärzte zu Herzinsuf.	Lebensqualität (QoL)	Arzt/Praxis	37/168
Rosemann 2007	Cox-/Gonarthrose	11: Ärzteschulung zu Arthrose 12: I1 + „Case-management“ durch MFAs	Keine Intervention	Lebensqualität (QoL)	Arzt/Praxis	75/1125
Szecsényi 2011	Diabetes mellitus Typ II	DMP-Behandlung + Praxisupport	DMP-Behandlung	HbA1c, Blutdruck	Arzt/Praxis	177/7807
Tinsel 2013	Hypertonie	Ärzteschulung shared-decision-making (SDM)	Keine Intervention	Patientenbeteiligung, Blutdruck	Arzt/Praxis	37/1120
Vollmar 2007	Demenz	Ärzteschulung zu evidenzbasierter Demenzbehandlung (2 Gruppen)	Standardtherapie	Zeit bis Eintritt in Pflegeheim oder Tod	Arzt/Praxis	303/390
Vollmar 2010	Hausärzte	Falldiskussionen + Schulungen zum Thema Demenz	Frontalunterricht zum Thema Demenz	Wissenszuwachs der Ärzte	QZ/CME	26/305
Vormfelde 2014	oral antikoagulierte Pat.	Patientenschulung durch MFAs	Informationsmaterial	Wissensstand, INR, Blutungskompl.	Arzt/Praxis	22/345

\* die erste Zahl beschreibt die Anzahl der randomisierten Cluster, die zweite Zahl die der rekrutierten Patienten/Probanden \*\* Hauptzielkriterium Sturzhäufigkeit noch nicht veröffentlicht

Abkürzungen (weitere Details zu den Abkürzungen finden sich im Abkürzungsverzeichnis):

Blutungskompl. = Blutungskomplikation(en); BMI = Body-Mass-Index; bzgl. = bezüglich; CME = continuing medical education; DMP = disease management program; FFbH-R = Funktionsfragebogen Hannover – Rückenschmerz-Version; HbA1c = glykiertes Hämoglobin; I1/I2 = Interventionsgruppe 1/2; INR = international normalized ratio; KHK = koronare Herzkrankheit; MFA = medizinische Fachangestellte; Pat. = Patient(en); QZ = Qualitätszirkel; Random. = Ebene, auf der die Randomisierung stattgefunden hat; vs. = versus; Wo. = Woche(n)

Tabelle 20: Abgeschlossene RCTs – „klassisch“ randomisiert (n = 16) Für Zuordnung von Studien zu Publikationen siehe „Übersicht aller Studienreferenzen“ im Literaturverzeichnis.

Erstautor Jahr	Teilnehmer/Erkrankung	Intervention(en)	Vergleichsgruppe	Zentrale Zielkriterien	Random.	Fallzahl*
<b>Studien zu definierten Behandlungen</b>						
Bleidorn 2010	Harnwegsinfekt	Ibuprofen	Ciprofloxacin	Symptome am Tag 4	Pat./Prob.	80
Bücker 2010	Rückenschmerz	Patientengespräch + schriftliche Infos zum Thema Rückenschmerz	Wie IG, aber Infos zu Alternativthema	Funktionskapazität	Pat./Prob.	189
Du Moulin 2009	COPD	Trainingsprogramm	Keine Intervention	6-Minuten Gehstest, LuFu, QoL	Pat./Prob.	20
Frese 2012	Hausarztpat. > 70 Jahre	Geriatrisches Assessment	Keine Intervention	Mortalität, Eintritt Altenheim	Pat./Prob.	1620
Gastpar 2003	Angst, Ruhelosigkeit	Kava Spezialextrakt WS 1490**	Placebo	Anxiety Status Inventory (ASI)	Pat./Prob.	141
Hensler 2009	Erkältung	Eigenbluttherapie	Placebo	Krankheitsdauer	Pat./Prob.	139
Jobst 2005	Häufige Atemwegsinfekte	i. m.-Eigenbluttherapie	i. m.-Engystol@**	Krankheitsdauer	Pat./Prob.	80
Klein 2013	Nackenschmerzen	Osteopathische strain-counterstrain Behandlung	Placebo-Behandlung	Bewegungsumfang + Schmerzintens.	Pat./Prob.	61
Peters-Klimm 2010	Herzinsuffizienz	„Case-management“ durch MFAs	Standardtherapie	Lebensqualität (QoL)	Pat./Prob.	199
Schencking 2013	Cox-/Gonarthrose	I1: Kneipp Hydrotherapie I2: Physiotherapie	I1 und I2	Schmerzintensität, Lebensqualität (GoL), Mobilität	Pat./Prob.	30
Voigt 2011	Migräne	Osteopathische Behandlung	Standardtherapie	Migraine Disability Assessment Score (MIDAS), GoL, Schmerzintensität	Pat./Prob.	42
<b>andere Studien</b>						
Bergold 2013	Ärztliche Berufsanfänger	Online-Kurs zum Thema Evidenzbasierte Medizin	Warteliste	Wissenszuwachs der Ärzte	Pat./Prob.	120
Blank 2013	Medizinstudenten	Untersuchungskurs + Peerteaching (durch PJ-Studenten)	Nur Untersuchungskurs	Objective structured clinical examination (OSCE)	Pat./Prob.	84
Butzlaff 2004	Hausärzte	Zugang zu computerbasierten Leitlinien	Keine Intervention	Wissenszuwachs der Ärzte	Pat./Prob.	72
Hoffmann 2014	Ärzte und MFAs allgemeinmedizinischer Praxen	Intensive Schulung des gesamten Praxisteam zu Patientensicherheit und Fehlermanagement	Fakultativer Vortrag zu Patientensicherheit und Fehlermanagement	Fehlermanagement gemessen mittels patient safety culture indicators (PSCIs)	Arzt/Praxis	65
Müller-Bühl 2011	Pat. mit ≥ 1 chronischen Erkrankung(en)	Beantwortung von 3 Übungsfragen vor dem Ausfüllen des SF-36 Fragebogens	Reguläres Ausfüllen des SF-36 Fragebogens	Anzahl an unbeantworteten Fragen	Pat./Prob.	215

\* die Fallzahl ist die Anzahl der Patienten/Probanden, die an der entsprechenden Studie teilgenommen haben und in die statistische Auswertung der Studie miteingegangen sind

\*\* hierbei handelt es sich um die Eigennamen eines naturheilkundlichen (Kava Spezialextrakt) bzw. eines homöopathischen Arzneimittels (Engystol®)

Abkürzungen (weitere Details zu den Abkürzungen finden sich im Abkürzungsverzeichnis):

COPD = chronic obstructive pulmonary disease; I1/I2 = Interventionsgruppe; i. m. = intramuskulär; LuFu = Lungenfunktionsprüfung; MFA = medizinische Fachangestellte; Pat. = Patient(en); PJ = Praktisches Jahr; Prob. = Proband(en); GoL = quality of life; Random. = Ebene, auf der die Randomisierung stattgefunden hat; Schmerzintens. = Schmerzintensität; SF-36 = short form 36 (Fragebogen zur Lebensqualität); vs. = versus; Wo. = Woche

Tabelle 21: Studienprotokolle ohne Ergebnisse (n = 17)

Für Zuordnung von Studien zu Publikationen siehe „Übersicht aller Studienreferenzen“ im Literaturverzeichnis.

Erstautor Jahr	Teilnehmer/Erkrankung	Intervention(en)	Vergleichsgruppe	Zentrale Zielkriterien	RCT-Art*	Fallzahl**
<b>Studien zu „prozessorientierten“ Interventionen</b>						
Altiner 2012a	Atemwegsinfektion	I1: Ärzteschulung I2: I1 + point-of-care Diagnostik	Standardtherapie	Anz. verschriebener Antibiotika	Cluster	188/13160
Altiner 2012b	Pat. mit $\geq 3$ chron. Erkrankungen	Multimodale Patientenschulung	Standardtherapie	Medikamenteneinnahme, QoL	Cluster	48/600
Barzel 2013	Pat. mit Armschwäche nach Apoplex	Modifizierte CIMT	Standardtherapie	Arm- u. Handfertigkeiten, WMFT	Cluster	48/144
Bozorgmehr 2014	D. m. II + $\geq 2$ chron. Erkrankungen	MFA-Hausbesuche + tel.-Anb. + BZ-Messung	Standardtherapie	Selbstfürsorge, HbA1c, QoL	„klassisch“	582
Drewelow 2012	Diabetes mellitus Typ II	Ärzteschulung zum Thema patientenzentrierte Kommunikation	Keine Intervention	HbA1c, Lebensqualität (QoL), Patientenpartizipation	Cluster	60/780
Freund 2011	COPD, Diabetes mellitus Typ II, Herzinsuffizienz	„Case-management“ durch MFAs + schriftliche Patienteninformation	Standardtherapie	Hospitalisationsrate, QoL, Versorgungsqualität	Cluster	130/2210
Gummersbach 2013	Frauen (48–49 Jahre)	Neue Informationsbroschüre zum Thema Mammographiescreening	Alte Broschüre	Intention zur Teilnahme am Mammographiescreening	„klassisch“	346
Jäger 2013	Multimorbide Pat. mit $\geq 4$ Medikamenten	Schulung für Qualitätszirkel (QZ) zu Medikamentenmanagement	Keine Schulung	Ausmaß der Umsetzung von Verbesserungsvorschlägen	Cluster	4/6800
Löffler 2014a	Niedergelassene Zahnärzte (ZÄ)	Schulung für ZÄ zum Einsatz von Antibiotika	Keine Intervention	Anz. verschriebener Antibiotika	Cluster	58/46000
Löffler 2014b	Geriatrische Krankenhauspatienten	Interv. zur Reduktion von Polypharmakotherapie	Keine Intervention	QoL, Anzahl Medikamente	Cluster	42/1626
Mortsiefer 2008	Hypertonie	Leitlinie „kardiovaskuläres Risikomanagement“ + Nur Versand der Leitlinie an Praxen Praxisbesuch	Nur Versand der Leitlinie an Praxen	kardiovaskuläres Risiko, einzelne Risikofaktoren	Cluster	92/3146
Schmidt 2010	Rückenschmerz	Pat.-Schulung (abh. vom Chronifizierungsrisiko)	Standardtherapie	Funktionskap., Krankheitstage	Cluster	40/600
Schmidt 2014	Pat. mit Zustand nach Sepsis	Entlassungsmanagement, Pat.- und Ärzteschulung, telefonische Anbindung	Standardtherapie	Lebensqualität (QoL), sepsisbezogene Spätfolgen	„klassisch“	290
Siebenhofer 2012	Pat. mit Indikation zur oralen Antikoagulation	Multimodale Intervention (Schulung, „Case-management“ etc.)	Standardtherapie	Anzahl Thrombembolien und Blutungskomplikationen	Cluster	46/690
Zimmermann 2014	Angsterkrankung, Depression oder somatoforme Störung	„Case-management“ und Beratung durch MFAs	Standardtherapie	Selbstwirksamkeit, Ausmaß psychiatrischer Symptome	Cluster	20/340
<b>Studien zu definierten Behandlungen</b>						
Gágyor 2012	Unkomplizierter Harnwegsinfekt	Initiale Therapie mit Ibuprofen, später bei Bedarf auch antibiotisch	Sofortige antibiotische Therapie	Symptomswere, Anzahl verschriebener Antibiotika	„klassisch“	494
Gensichen 2014	Panikstörung mit/ohne Agoraphobie	Psychoedukation, Expositionsübungen, tel.-Anb.	Standardtherapie	Schwere der Angststörung	Cluster	74/444

\* RCT-Art = Art der Randomisierung (cluster oder „klassisch“) \*\* Fallzahl = geplante Anzahl an Studienteilnehmern; bei cRCTs steht an erster Stelle die Anzahl der geplanten Cluster

Abkürzungen (weitere Details zu den Abkürzungen finden sich im Abkürzungsverzeichnis):

abh. = abhängig; Anz. = Anzahl; BZ = Blutzucker; chron. = chronisch; CIMT = constraint-induced movement therapy; COPD = chronic obstructive pulmonary disease; D. m. II = Diabetes mellitus Typ II; etc. = et cetera; Funktionskap. = Funktionskapazität; HbA1c = glykiertes Hämoglobin; I1/I2 = Interventionsgruppe 1/2; Interv. = Intervention; MFA = medizinische Fachangestellte; Pat. = Patient(en); QoL = quality of life; tel.-Anb. = telefonische Anbindung; u. = und; WMFT = Wolf Motor Function Test

### 3.2.5 Informationen zum Studiendesign

Von den 50 eingeschlossenen Studien waren 30 (60 %) als clusterrandomisierte Studien konzipiert. Während unter den 33 abgeschlossenen RCTs (mit Ergebnissen) bei 17 (52 %) eine Clusterrandomisierung vorlag, war für 13 (76 %) der 17 RCTs ohne Ergebnisse (nur Studienprotokolle) eine Clusterrandomisierung vorgesehen. (Tabelle 22)

Tabelle 22: Randomisierungsart der eingeschlossenen RCTs

Art der Randomisierung	RCTs mit Ergebnissen n = 33	RCTs ohne Ergebnisse (Studienprotokolle) n = 17	Gesamt n = 50
clusterrandomisiert	17	13	30 (60 %)
„klassisch“ randomisiert	16	4	20 (40 %)

Die Analyse der folgenden Studienmerkmale erfolgt nicht – wie in den bisherigen Tabellen – getrennt nach Studien mit Ergebnissen und Studien ohne Ergebnisse (nur Studienprotokolle), sondern aufgeteilt nach clusterrandomisierten und „klassischen“ randomisierten Studien. Der Grund hierfür ist, dass die Gegenüberstellung von clusterrandomisierten und „klassischen“ randomisierten Studien eine bessere Hervorhebung der innerhalb dieser Studiengruppen vorherrschenden speziellen Charakteristika und Problemfelder ermöglicht. Die im Folgenden genannten Angaben beziehen sich im Falle von abgeschlossenen Studien auf die in den zugehörigen Publikationen dokumentierten Ergebnisse, im Falle der Studien ohne Ergebnisse auf die in den Studienprotokollen genannten Planungen.

Von den 50 untersuchten randomisierten Studien wiesen 45 (90 %) zwei Studienarme auf, in 5 Studien (10 %) gab es drei Studienarme (Tabelle 23). Der Anteil an dreiarmligen Studien war unter den clusterrandomisierten Studien mit 4 (13 %) von 30 höher als bei den „klassischen“ randomisierten Studien mit 1 (5 %) von 20.

Im Allgemeinen kann gesagt werden, dass bei den 30 clusterrandomisierten Studien die Randomisierungsebene (engl. unit of randomisation) meist aus den beteiligten Ärzten oder Praxen bestand. Handelte es sich bei den beteiligten Praxen ausschließlich um Einzelpraxen, so war eine Unterscheidung der Randomisierungsebene in Arzt bzw. Praxis (da identisch) nicht nötig. Waren hingegen Gemeinschaftspraxen beteiligt, so war es durchaus von Interesse, ob diese Praxen (mit allen dort tätigen Ärzten) „als Ganzes“ randomisiert wurden oder ob die Randomisierung für alle in der Gemeinschaftspraxis

tätigen Ärzte einzeln erfolgte (in diesem Fall wäre es möglich gewesen, dass die verschiedenen Ärzte einer Praxis unterschiedlichen Gruppen zugeordnet sind). Da jedoch für viele Studien nicht eindeutig nachvollziehbar war, ob Ärzte oder Praxen randomisiert wurden („Ärzte“ bzw. „GPs“ und „Praxen“ bzw. „general practices“ wurden häufig synonym verwendet) und die zahlenmäßigen Unterschiede zwischen beteiligten Ärzten und Praxen aufgrund eines hohen Anteils an Einzelpraxen sehr gering war (Anhang D), wird zwischen ihnen im Weiteren nicht mehr unterschieden:

Die Randomisierungsebene bestand bei 25 (83 %) der 30 clusterrandomisierten Studien aus beteiligten Ärzten oder Praxen (in Einzelfällen waren auch andere Berufsgruppen wie z. B. Physiotherapeuten oder Zahnärzte bzw. deren Praxen beteiligt). In 4 Fällen bestanden die Cluster aus Qualitätszirkeln bzw. „CME-Groups“ [Jäger et al., 2013, Kaufmann-Kolle et al., 2011, Krones et al., 2008, Vollmar et al., 2010a] und in einem Fall aus Krankenhausstationen [Löffler et al., 2014b].

Bei 15 (75 %) der 20 „klassischen“ RCTs erfolgte die Randomisation auf der Ebene der Patienten/Probanden. In jeweils einem Fall wurden Medizinstudenten, ärztliche Berufsanfänger, gesamte Praxisteams, Praxen und Paare (bestehend aus Patient und behandelndem Hausarzt) randomisiert.

Tabelle 19 und Tabelle 20 (PICO-Tabellen) können die Randomisierungsebenen der 33 abgeschlossenen RCTs (mit Ergebnissen) entnommen werden.

Ein (im Vorhinein) abgeschätzter ICC war für 26 der 30 clusterrandomisierten Studien genannt. Dieser betrug minimal 0,01, maximal 0,2.

Ein Intracluster-Korrelationskoeffizient (engl. intracluster correlation coefficient, ICC) wurde in den Publikationen zu 6 (35 %) der 17 clusterrandomisierten Studien mit Ergebnissen (Tabelle 22) berichtet:

- Altiner et al. [2007]: ICC = 0,2
- Erler et al. [2012]: ICC = 0,039
- Peters-Klimm et al. [2009]: ICC = „fast 0“
- Vollmar et al. [2007b]: ICC = 0,03
- Vollmar et al. [2010a]: ICC = 0,054
- Vormfelde et al. [2014]: ICC = 0,18

Für 15 (88 %) der 17 clusterrandomisierten Studien mit Ergebnissen (Tabelle 22) war angegeben, dass die Clusterstruktur bei der Datenauswertung (z. B. durch die Verwendung spezieller statistischer Analysemodelle wie „generalized estimating equations“, „mixed models“, „generalized linear models“ oder „multi-level models“) berücksichtigt wurde, bei

10 der 13 clusterrandomisierten Studien ohne Ergebnisse war dies zumindest angekündigt.

Für 42 (84 %) der 50 eingeschlossenen RCTs waren Fallzahlschätzungen (engl. sample size estimations) berichtet. Sie bezogen sich auf 29 (97 %) der 30 clusterrandomisierten und 13 (65 %) der 20 „klassischen“ randomisierten Studien. Allerdings kann davon ausgegangen werden, dass in manchen Publikationen eine erfolgte Fallzahlschätzung nicht berichtet wurde. Dies war vermutlich vor allem dann der Fall, wenn in einer Publikation bereits Studienergebnisse veröffentlicht wurden, sodass die geplanten Fallzahlen einer im Vorhinein erfolgten Fallzahlschätzung zugunsten der „interessanteren“ tatsächlichen Fallzahlen keine Erwähnung mehr fand. Somit war in Wirklichkeit wahrscheinlich sogar bei einem höheren Anteil als dem o. g. eine Fallzahlschätzung erfolgt. Diese Vermutung wird auch dadurch gestützt, dass – bei einer getrennten Betrachtung der 17 Studienprotokolle (ohne Ergebnisse) und der 33 Studien mit Ergebnissen – in *allen* 17 (100 %) Studienprotokollen (ohne Ergebnisse) eine Fallzahlschätzung beschrieben war, bei den Studien mit Ergebnissen hingegen nur in 25 (76 %) von 33 Fällen. In Tabelle 21 finden sich die geschätzten Fallzahlen der 17 Studienprotokolle (ohne Ergebnisse) aufgeführt.

Eine „Intention-to-treat-Analyse“ (ITT-Analyse) war für 28 (56%) der 50 eingeschlossenen Studien durchgeführt oder angekündigt worden, während dies für 16 (32 %) Studien nicht der Fall war. Für die verbleibenden 6 (12 %) Fälle konnte nicht genau beurteilt werden, ob eine entsprechende Analyse durchgeführt worden war. Bei der getrennten Betrachtung der Randomisierungsarten zeigte sich, dass 18 (60 %) der 30 clusterrandomisierten und 10 (50 %) der 20 „klassischen“ randomisierten Studien von einer ITT-Analyse berichteten.

Die Studiendauer, also die Dauer vom Einschluss eines Patienten bzw. Probanden in die Studie bis zum Ende der Studienteilnahme dieser Person, wurde ebenfalls erfasst. Bei zwei Studien [Klein et al., 2013, Müller-Bühl et al., 2011] war aufgrund des Studiendesigns von einer Studiendauer von unter einer Stunde auszugehen. Frese et al. [2012] wiesen mit bis zu 7 Jahren die längste Studiendauer unter den untersuchten Studien auf.

Tabelle 23 zeigt auf, dass die Studiendauer bei 22 (44 %) der 50 eingeschlossenen Arbeiten unter 6 Monaten lag. Der Anteil der bis zu 6 Monate dauernden Studien war bei den „klassischen“ randomisierten Studien (12 von 20 Studien, 60 %) fast doppelt so hoch wie bei den clusterrandomisierten Studien (10 von 30 Studien, 33 %). Über ein Drittel (11 von 30 Studien, 37 %) der clusterrandomisierten Studien hatte eine Studiendauer von mehr als 6, aber maximal 12 Monaten. Bei immerhin fast einem Viertel (7 von 30 Studien, 23 %) der clusterrandomisierten Studien betrug die Studiendauer mehr als 18, aber maximal 24 Monate.

Somit kann gesagt werden, dass die Stundendauer bei den cRCTs im Vergleich zu den „klassischen“ RCTs tendenziell länger war. Dies zeigt sich auch am jeweiligen Median der Studiendauer in Tagen für die „klassischen“ (Median = 140 Tage) und die clusterrandomisierten (Median = 360 Tage) Studien.

Das Ausmaß und die Art der Verblindung der eingeschlossenen Studien wurden untersucht, indem für jede Studie die folgenden drei Fragen beantwortet wurden (2.5.5):

- Waren die Patienten/Probanden verblindet?
- Waren die Ärzte/Behandler verblindet?
- War die „Outcomemessung“ verblindet?

Von den 50 eingeschlossenen Studien wiesen 37 (74 %) *keinerlei* Verblindung auf, d. h. alle drei Fragen konnten mit „nein“ beantwortet werden. 4 (8 %) der Studien waren „vollständig“ verblindet, d. h. alle drei Fragen konnten mit „ja“ beantwortet werden (dies wird auch häufig als „doppelblind“ bezeichnet). Bei 5 (10 %) Studien war ausschließlich die „Outcomemessung“ verblindet. Bei den 4 (8 %) verbleibenden Studien lagen andere Formen der teilweisen Verblindung vor. (Tabelle 23)

Betrachtet man die Verblindung getrennt nach cluster- und „klassischen“ randomisierten Studien, so zeigt sich, dass 27 (90 %) der 30 clusterrandomisierten und 10 (50 %) der 20 „klassischen“ randomisierten Studien *keinerlei* Verblindung aufwiesen. Während *keine* (0 %) der clusterrandomisierten Studien „vollständig“ verblindet war, traf dies auf 4 (20 %) der 20 „klassischen“ randomisierten Studien zu.

Bei der Analyse der 50 eingeschlossenen Studien sollte auch die sogenannte Interventionsebene untersucht werden (2.5.5). Es erfolgte also eine Einteilung danach, ob die Intervention primär am Arzt (z. B. in Form einer Fortbildung für Ärzte zum Thema Demenz [Vollmar et al., 2007b]), am Praxisteam (z. B. in Form einer Schulung für MFAs, die sie zur Durchführung eines „Case-managements“ bei Patienten mit depressiver Störung befähigt [Gensichen et al., 2009]) oder direkt am Patienten (z. B. in Form einer „Medikamentenstudie“, bei der Patientinnen mit Harnwegsinfekt entweder ein Antibiotikum oder ein Schmerzmittel erhalten [Gágyor et al., 2012]) ansetzte.

Diese Abgrenzung war allerdings schwierig, da sich eine Intervention auf der Ebene der Ärzte bzw. der Praxisteams (z. B. in Form einer Schulung zur Implementierung eines „Case-management“-Programms in der Hausarztpraxis) im weiteren Verlauf häufig direkt auf die Patienten auswirkte und somit auch als „Intervention am Patienten“ gewertet werden konnte (z. B. indem Ärzte bzw. Praxisteams das erlernte „Case-management“ am Patienten umsetzten).

Deshalb wurde für jede Studie zunächst die *primäre* Interventionsebene erfasst (also die Ebene bzw. Personengruppe, an der die Intervention *unmittelbar* ansetzte). Von den 50 untersuchten Studien war dies bei 10 (20 %) Studien die ärztliche Ebene und bei 9 (18 %) Studien die Ebene der MFAs bzw. der Praxisteams oder anderer Studienassistenten. Bei 15 (30 %) Studien zielte die untersuchte Intervention direkt auf den Patienten/Probanden ab. In 5 (10 %) Studien bezog sich die Interventionen zwar auf Ärzte, Medizinstudenten oder das Praxisteam, allerdings stellten diese selbst die Studienobjekte dar, ohne dass (andere) Patienten/Probanden an der Studie teilgenommen haben (z. B. Studien zum Wissenszuwachs von Ärzten bei verschiedener Lehr-Interventionen [Bergold et al., 2013]). Bei 11 (22 %) Studien lag eine Kombination der Interventionsebenen vor.

Betrachtet man die Interventionsebenen separat für clusterrandomisierte und „klassische“ randomisierte Studien, so zeigt sich, dass bei den clusterrandomisierten Studien in 10 (33 %) von 30 Fällen eine kombinierte Intervention auf verschiedenen Ebenen und ebenfalls in 10 (33 %) von 30 Fällen eine Intervention auf ärztlicher Ebene vorlagen. Bei den „klassischen“ randomisierten Studien setzten die Interventionen hingegen in 12 (60 %) von 20 Fällen auf Patientenebene an.

Darüber hinaus wurde untersucht, ob die Interventionen – unabhängig von deren primären Ansatzpunkten – tatsächlich „bei den Patienten ankamen“. Dies war bei 5 (10 %) der 50 Studien eindeutig nicht der Fall, da Ärzte/Medizinstudenten selbst (z. B. im Rahmen von Studien zur Lehre) die Probanden darstellten und keinerlei Patienten an diesen Studien teilnahmen. In 12 (24 %) der 50 Studien kamen die Patienten nur mittelbar mit der Intervention „in Berührung“ (Beispiel: Im Rahmen einer cRCT erhielten Ärzte der Interventionsgruppe eine Schulung zum Thema „shared decision making“ [Krones et al., 2008]). Auch wenn diese Schulung vermutlich keine für die Patienten „spürbare“ Intervention darstellte, so ist trotzdem davon auszugehen, dass sie einen Einfluss auf die Arzt-Patienten-Kommunikation hatte. In welchem Ausmaß dies jedoch der Fall war, ist schwer beurteilbar. Deshalb wird die Intervention als „sich nur mittelbar auf die Patienten auswirkend“ eingestuft.). In 33 (66 %) der 50 Studien „kam“ die Intervention direkt bei den Patienten „an“. Die Patienten erhielten also z. B. ein „Case-management“ mit regelmäßigen Telefonkontakten [Gensichen et al., 2009], Physiotherapie [Schencking et al., 2009] oder Ähnliches.

### 3 Ergebnisse

Tabelle 23: Studiencharakteristika, gruppiert nach der Art der Randomisierung

	Cluster- RCTs n = 30	„klassische“ RCTs n = 20	Gesamt n = 50
<b>Studienarme</b>			
- zwei	26	19	45 (90 %)
- drei	4	1	5 (10 %)
<b>Studiendauer (d) pro Proband von Studieneinschluss bis –ende</b>			
- 24 Mon. < d	1	1	2 (4 %)
- 18 Mon. < d ≤ 24 Mon.	7	1	8 (16 %)
- 12 Mon. < d ≤ 18 Mon.	1	1	2 (4 %)
- 6 Mon. < d ≤ 12 Mon.	11	4	15 (30 %)
- d ≤ 6 Mon.	10	12	22 (44 %)
- nicht angegeben	-	1	1 (2 %)
- Studiendauer/Proband (in Tagen)*	360 (14–1460)	140 (1–2263)	270 (1–2263)
<b>Verblindung</b>			
- „keinerlei“	27	10	37 (74 %)
- nur „Outcomemessung“ verblindet	1	4	5 (10 %)
- „vollständig“	-	4	4 (8 %)
- andere Formen der teilweisen Verblindung	2	2	4 (8 %)
<b>primäre Interventionsebene</b>			
- Arzt	10	-	10 (20 %)
- MFA/Team	6	3	9 (18 %)
- Patient/Proband	3	12	15 (30 %)
- Proband = Arzt, Medizinstudent, Praxisteam	1	4	5 (10 %)
- Kombination	10	1	11 (22 %)
<b>„Kommt die Intervention am Patienten an?“</b>			
- ja	17	16	33 (66 %)
- nein	1	4	5 (10 %)
- nur mittelbar	12	-	12 (24 %)

\* hierbei handelt es sich um die Zeit zwischen dem Studieneinschluss eines Probanden und der letzten Datenerhebung zu ebendiesem, angegeben in Tagen [Median (Minimum–Maximum)]

Abkürzungen (weitere Details zu den Abkürzungen finden sich im Abkürzungsverzeichnis):

d = Studiendauer; MFA = Medizinische Fachangestellte

Unter 3.2.4.1 und in Tabelle 15 wurde bereits beschrieben, dass sich 31 (62 %) der 50 eingeschlossenen Studien mit „prozessorientierten“ Interventionen befassten, während 14 (28 %) RCTs Interventionen zu „einer definierten Behandlung“ untersuchten. In 3.2.4.1 wurden dabei die Interventionsarten („prozessorientiert“ und „definierte“ Behandlung) dahingehend untersucht, ob die Studien bereits abgeschlossen waren oder sich noch in Planung/Durchführung (Studienprotokolle ohne Ergebnisse) befanden. Zusätzlich erfolgte allerdings eine Analyse der Interventionsarten getrennt nach der Art der Randomisierung („klassisch“ vs. cluster). Sie ergab, dass 25 (83 %) der 30 clusterrandomisierten Studien „prozessorientierte“ Interventionen thematisierten, während 10 (50 %) der 20 „klassischen“ RCTs Interventionen zu „definierten“ Behandlungen untersuchten.

Tabelle 24: Einteilung der eingeschlossenen Studien nach Interventions- und Randomisierungsart

Interventionsart	Cluster-RCTs n = 30	„klassische“ RCTs n = 20	Gesamt n = 50
„definierte Behandlung“	4	10	14 (28 %)
„prozessorientierte“ Behandlung	25	6	31 (62 %)
andere	1	4	5 (10 %)

### 3.2.6 Orte der Studiendurchführung

Von den 50 eingeschlossenen Studien fanden 9 (18 %) Studien an einem einzigen Studienzentrum statt, während 41 (82 %) an mehreren Studienzentren (multizentrisch), z. B. in verschiedenen Arztpraxen, durchgeführt wurden (Tabelle 25).

Bei denjenigen 9 Studien, die nur an einem Studienzentrum stattfanden, konnte in 6 Fällen nicht ermittelt werden, um welche Art von Einrichtung es sich bei diesem Studienzentrum handelte (hier wurde lediglich aus dem Kontext heraus klar, dass sie an *einem* Ort stattgefunden hatten). Jeweils in einem Fall handelte es sich um eine Kurklinik [Schencking et al., 2013], um eine große Hausarztpraxis, die aber der alleinige Durchführungsort dieser RCT war [Klein et al., 2013], und um die medizinische Fakultät einer Universität [Blank et al., 2013].

Von den 41 multizentrischen Studien fanden 39 in Praxen (in 37 Fällen handelte es sich dabei um „normale“ hausärztliche Praxen, in jeweils einem Fall um physiotherapeutische [Barzel et al., 2013] und zahnärztliche [Löffler et al., 2014a] Praxen) und 2 auf Stationen unterschiedlicher Krankenhäuser [Löffler et al., 2014b, Schmidt et al., 2014] statt.

### 3 Ergebnisse

Somit waren 78 % aller in der vorliegenden Arbeit untersuchten Studien in (mehreren) Praxen lokalisiert. Unter den clusterrandomisierten Studien lag der Anteil derjenigen Studien, die in (mehreren) Praxen durchgeführt wurden sogar bei 93 %, während der entsprechende Wert bei den „klassischen“ randomisierten Studien 55 % betrug.

Tabelle 25: Anzahl Studienzentren und Durchführungsorte der eingeschlossenen Studien

Anzahl und Art der Studienzentren	RCTs mit Ergebnissen (n = 33)	RCTs ohne Ergebnisse (Studienprotokolle) (n = 17)	Gesamt (n = 50)	Cluster-RCTs (n = 30)	„klassische“ RCTs (n = 20)
<b>ein Studienzentrum</b>					
Art von Studienzentrum unbekannt	6	-	6 (12 %)	1	5
Kurklinik	1	-	1 (2 %)	-	1
große Hausarztpraxis	1	-	1 (2 %)	-	1
Universität	1	-	1 (2 %)	-	1
<b>mehrere Studienzentren</b>					
Arztpraxen	24	13	37 (74 %)	26	11
Praxen anderer Heilberufe	-	2	2 (4 %)	2	-
Krankenhausstationen	-	2	2 (4 %)	1	1

#### 3.2.7 Beteiligte und Teilnehmer der untersuchten Studien

Im Folgenden finden sich – soweit bei teilweise lückenhaften Angaben in den untersuchten Publikationen möglich – Informationen zu „Beteiligten“ und „Teilnehmern“ der untersuchten Studien.<sup>1</sup> Es werden dabei deren Rekrutierungsprozesse und Anzahlen sowie deren Arbeitsaufwand und die ihnen gewährten Teilnahmeanreize (engl. incentives) qualitativ beschrieben.

<sup>1</sup> In der vorliegenden Arbeit werden Ärzte, Praxen und Assistenzteams, die an den jeweiligen Studien mitgewirkt haben, als „Studienbeteiligte“ bezeichnet. Patienten/Probanden werden hingegen „Studienteilnehmer“ genannt.

### 3.2.7.1 Rekrutierung

Zur Rekrutierung von „Studienbeteiligten“ wurden in der Regel zunächst Kontaktdaten von in Frage kommenden Ärzten bzw. Praxen gesammelt. Diese stammten in den meisten Fällen von der zuständigen Kassenärztlichen Vereinigung. Seltener wurde auf das Lehrpraxen-Netzwerk von Universitäten oder auf ärztliche Qualitätszirkel zurückgegriffen. Anschließend erfolgte meist mittels dieser Kontaktdaten die (häufig schriftliche) Einladung zur Beteiligung an der Studie.

Bei den „Studienteilnehmern“ (Patienten/Probanden) erfolgte die Rekrutierung meist konsekutiv (z. B. wurden in vielen Fällen alle in die Praxis kommenden Patienten, welche die Einschlusskriterien erfüllten, nacheinander um ihre Studienteilnahme gebeten) oder über die Auswahl per Computersoftware. In seltenen Fällen wurde öffentlich für die Studienteilnahme geworben (z. B. mittels Zeitungsinseraten, Flyern oder Aushängen in der Praxis) oder es wurden andere Verfahren (wie die individuelle Patientenauswahl durch den Arzt) gewählt.

### 3.2.7.2 Anzahl

Für die 39 in Praxen stattfindenden multizentrischen Studien (Tabelle 25) sollte die Anzahl beteiligter Ärzte<sup>1</sup> und Praxen (die Anzahl von Ärzten und Praxen wich voneinander ab, wenn in mindestens einer beteiligten Praxis mehr als ein Arzt tätig war) zu Studienbeginn ( $t_0$ ) und zum Zeitpunkt der ersten Datenerhebung ( $t_1$ ) ermittelt werden. Die Erhebung dieser Informationen gestaltete sich jedoch schwierig, da häufig nur die Anzahl von Ärzten *oder* die Anzahl von Praxen genannt war. Deshalb entschied sich der Autor der vorliegenden Arbeit zur Entwicklung eines Schätzungsinstruments (Anhang D), anhand dessen für jede Studie die fehlende Anzahl (an Ärzten oder Praxen) bei Kenntnis der jeweils anderen Anzahl annäherungsweise bestimmt werden konnte.

Im Folgenden wird nicht zwischen tatsächlich erhobenen und mittels des o. g. Schätzungsinstruments bestimmten Zahlen unterschieden. Zudem wird aus Gründen der besseren Übersichtlichkeit nur auf die beteiligten Ärzte, nicht jedoch auf die beteiligten Praxen eingegangen, da eine tiefere Differenzierung aufgrund der nur geringfügig voneinander abweichenden Werte an dieser Stelle nicht notwendig ist. Detailinformationen können Tabelle 33 in Anhang D entnommen werden.

---

<sup>1</sup> Hierbei handelte es sich in jeweils einer Studie um Zahnarzt- bzw. physiotherapeutische Praxen; der Einfachheit halber wird aber weiterhin von „Ärzten“ gesprochen.

Die mediane Anzahl beteiligter Ärzte unter den 33 Studien mit Ergebnissen betrug zum Zeitpunkt des Studienbeginns 41, die Spannweite war mit minimal 10 bis maximal 303 Ärzten groß. Mit einer medianen Anzahl von 40 beteiligten Ärzten zum Zeitpunkt der ersten Datenerhebung war die Abbrecherquote gering. In wenigen Fällen lagen die Abbrecherquoten unter den beteiligten Ärzten aber auch bei über 50 % [Mehring et al., 2014, Vollmar et al., 2007b].

Die mediane Anzahl der (geplanten) beteiligten Ärzte bei den 17 Studien ohne Ergebnisse (nur Studienprotokolle) lag mit 52 über der entsprechenden Anzahl von 41 bei den bereits abgeschlossenen Studien mit Ergebnissen.

Beim Vergleich der beiden in der vorliegenden Arbeit untersuchten Arten von randomisierten Studien („klassisch“ vs. cluster) zeigte sich, dass die mediane Anzahl der beteiligten Ärzte bei den 30 clusterrandomisierten Studien die vergleichbare mediane Anzahl bei den 20 „klassischen“ randomisierten Studien zu Studienbeginn ca. um den Faktor 3 überstieg (Tabelle 33 in Anhang D).

Bei 45 (90 %) der 50 eingeschlossenen Studien stellten „reguläre“ Probanden bzw. Patienten die Studienobjekte dar. Die initiale Anzahl der an diesen 45 Studien teilnehmenden Patienten/Probanden betrug median 390 (minimal 20, maximal 46 000). Auffallend ist die hohe Patientenzahl von 46 000, die in einem Studienprotokoll der DREAM-Studie [Löffler et al., 2014a] angegeben ist. Sie erklärt sich dadurch, dass bei dieser clusterrandomisierten Studie von 46 beteiligten Zahnarztpraxen ausgegangen wird, für die jeweils von 1 000 Patienten Daten erhoben werden sollen. Diese hohe Zahl stellt eine Ausnahme dar, die zweitgrößte Anzahl an Patienten beträgt 13 160 [Altiner et al., 2012a].

Unter den 33 abgeschlossenen Studien mit Studienergebnissen betrug die mediane Patientenzahl zu Studienbeginn 264,5 (minimal 20, maximal 7 807), während sie zum Zeitpunkt der ersten Datenerhebung bei 235,5 (minimal 20, maximal 7 245) lag. Die (geplante) mediane Patientenzahl zu Studienbeginn betrug bei den Studienprotokollen (ohne Ergebnisse) 690 (minimal 144, maximal 46 000).

Beim Vergleich der beiden in der vorliegenden Arbeit vorkommenden Arten von randomisierten Studien („klassisch“ vs. cluster) zeigte sich, dass die medianen Patientenzahlen bei den clusterrandomisierten Studien die entsprechenden Zahlen der „klassischen“ randomisierten Studien zu Studienbeginn um mehr als das 4-fache überstiegen.

Folgende Zahlen beziehen sich auf die verbleibenden 5 (10 %) Studien, bei denen – wie bereits unter 3.2.5 erwähnt – Ärzte, Medizinstudenten oder das Praxisteam selbst die Studienobjekte darstellten, sodass hier keine Patienten oder „reguläre“ Probanden mitwirkten. Diese Studien wurden unter 3.2.6 zu den unizentrischen Studien gezählt, da z. B. Praxen, deren Inhaber selbst als Studienprobanden fungieren, schlecht als „Studienzentren“ bezeichnet werden können. An diesen 5 Studien nahmen zu Studienbeginn im Median 84 Probanden (minimal 65, maximal 305) teil, während im Median 69 Probanden (minimal 53, maximal 166) die Studie abschlossen.

Tabelle 26: Anzahl der Patienten/Probanden zu Studienbeginn ( $t_0$ ) und zum Zeitpunkt der ersten Datenerhebung ( $t_1$ )\*

Anzahl teilnehmender Patienten/Probanden	RCTs mit Ergebnissen	RCTs ohne Ergebnisse (Studienprotokolle)	Gesamt	Cluster-RCTs	„klassische“ RCTs
initial ( $t_0$ )	264,5 (20–7807) n = 28	690 (144–46000) n = 17	390 (20–46000) n = 45	690 (144–46000) n = 29	165 (20–1620) n = 16
zum Zeitpunkt der ersten Datenerhebung ( $t_1$ )	235,5 (20–7245) n = 28	-	235,5 (20–7245) n = 28	390,5 (141–7245) n = 16	97 (20–1215) n = 12

\* dargestellt als Median (Minimum–Maximum)

### 3.2.7.3 Arbeitsaufwand

Bei einem großen Teil der eingeschlossenen Studien fanden sich keine Informationen über den zeitlichen Aufwand des beteiligten Studienpersonals oder der teilnehmenden Patienten. Leider können auch für diejenigen Studien, für die entsprechende Daten erhoben werden konnten, keine fundierten Aussagen zu dem Arbeitsaufwand der Beteiligten getroffen werden. Grund hierfür ist, dass diese Daten meist nur für einen kleinen Abschnitt der Studie (z. B. war die Dauer einer Schulung für Ärzte/Praxisteam genannt, während Informationen zum zeitlichen Aufwand der nachfolgenden Patientenbetreuung im Rahmen der Studie fehlten) oder nicht für alle Interventionsgruppen vorlagen.

### 3.2.7.4 „Incentives“

Für die Mehrheit der Studien lagen keine Informationen zu Mitwirkungsanreizen (engl. incentives) für Ärzte bzw. Praxisteams oder für Patienten/Probanden vor. Es konnte also nicht beurteilt werden, ob entsprechende Zahlungen erfolgt waren oder nicht. Doch selbst wenn für eine Studie bekannt war, dass Anreizzahlungen getätigt worden waren, ließ sich meist deren Höhe nicht eruieren oder die Angaben waren unvollständig (z. B. war angegeben, dass sowohl Ärzte als auch MFAs Aufwandsentschädigungen erhielten, allerdings wurde die Höhe derselben nur für eine der beiden Berufsgruppen genannt).

Trotz dieser Einschränkungen soll an dieser Stelle versucht werden einen groben Überblick über die Höhe der Aufwandsentschädigungen zu geben. Es sei jedoch auf die begrenzte Aussagekraft aufgrund der o. g. Problematik hingewiesen.

In 2 Studien erhielt jeder Arzt für seine Studienbeteiligung pauschal 200 € [Becker et al., 2008, die Informationen hierzu finden sich jedoch in Chenot et al., 2008] bzw. 300 € [Mortsiefer et al., 2008]. Bei den 6 weiteren Studien mit Angaben zur Höhe der Aufwandsentschädigungen wurden die Vergütungen für die beteiligten Ärzte pro teilnehmendem Patient gewährt. Dabei erhielten die Ärzte der Interventionsgruppe und die Ärzte der Kontrollgruppe in 3 Fällen mit jeweils 10 € [Tinsel et al., 2012], 90 € [Junius-Walker et al., 2012] und 100 € [Bleidorn et al., 2010, die Informationen hierzu finden sich jedoch in Gágyor et al., 2011] pro Patient eine Entschädigung in *gleicher* Höhe. In den anderen 3 Fällen unterschied sich die Vergütung je nach Gruppenzuordnung: Hier erhielten die Ärzte der Interventionsgruppe in 2 Studien [Mehring et al., 2013, Mehring et al., 2014] jeweils 50 € und in einer Studie [Gensichen et al., 2009, die Information hierzu findet sich jedoch in Güthlin et al., 2012] jeweils 100 € pro Patient, während den Ärzten in den Kontrollgruppen jeweils nur die Hälfte der Vergütung der Interventionsgruppen zukam.

In manchen Fällen wurden Sachleistungen gewährt, deren Wert sich nicht genau beziffern ließ. Z. B. durften Praxen einen für die Studie benötigten Tablet-PC [Jäger et al., 2013] oder ein Langzeit-Blutdruckmessgerät [Tinsel et al., 2012] nach Studienende behalten.

Informationen zu Incentives für Patienten/Probanden fanden sich nur in den wenigsten Fällen – meist blieb unklar, ob entsprechende Zahlungen erfolgt waren oder nicht. In 2 Studien [Mehring et al., 2013, Mehring et al., 2014] erhielten die Patienten der Interventionsgruppe einen kostenfreien Zugang zu einer Online-Plattform, welche die Patienten bei der Gewichtsreduktion bzw. Raucherentwöhnung unterstützen sollte (sonst wären diese Zugänge kostenpflichtig gewesen; z. B. kostet ein 12-wöchiger Zugang für das Online-Coaching zur Gewichtsreduktion 79 € [Mehring et al., 2013]). Die Patienten der Kontrollgruppen dieser beiden Studien erhielten jeweils 10 € für ihre Studienteilnahme.

Zu 2 weiteren Studien lagen Informationen zu „Incentives“ vor, allerdings handelte es sich hierbei – wie bereits unter 3.2.5 beschrieben – um Studien mit ärztlichen Probanden. In einer Studie [Vollmar et al., 2010a] wurde explizit darauf hingewiesen, dass die ärztlichen Probanden *keine* Anreizzahlungen für ihre Studienteilnahme erhalten hatten, in der anderen Studie [Kaufmann-Kolle et al., 2011] war angegeben, dass insgesamt 6 x 250 €, also in Summe 1 500 €, unter den Studienteilnehmern verlost wurden.

### 3.3 Verzerrungsrisiken der untersuchten Studien

Die Beurteilung der Verzerrungsrisiken der eingeschlossenen Studien (engl. Risk of bias assesment) erfolgte nur für die 33 bereits abgeschlossenen Studien mit Studienergebnissen.<sup>1</sup> Diese waren in 17 Fällen clusterrandomisiert, bei 16 lag eine „klassische“ Randomisation vor (Tabelle 22).

#### 3.3.1 Verzerrungsrisiken der clusterrandomisierten Studien

Bei vielen der 17 clusterrandomisierten Studien waren Methodik und logistischer Aufwand komplexer als bei den „klassischen“ randomisierten Studien. Dies führte im Durchschnitt zu (im Vergleich zu den „klassischen“ randomisierten Studien) höheren Verzerrungsrisiken.

Obwohl für keine der clusterrandomisierten Studien in Bezug auf die Erstellung der Randomisierungssequenz (engl. random sequence generation) die Bewertung „hohes Verzerrungsrisiko“ angenommen wurde (in 10 Studien war dieses Verzerrungsrisiko niedrig, in 7 Studien unklar), musste die Gefahr einer Stichprobenverzerrung (engl. selection bias) häufig als hoch eingestuft werden. Dies lag daran, dass die Geheimhaltung der Gruppenzuordnung (engl. allocation concealment) bei vielen Studien nicht gewährleistet war (in Bezug auf das „allocation concealment“ bestand bei 10 Studien ein hohes, bei 2 Studien ein unklares und in 5 Studien ein niedriges Verzerrungsrisiko).

Die häufig mangelnde „Geheimhaltung der Gruppenzuordnung“ ist partiell der Natur der clusterrandomisierten Studien geschuldet. Zwar ist eine „Geheimhaltung der Gruppenzuordnung“ theoretisch auch bei clusterrandomisierten Studien möglich, sie gestaltet sich aber in der praktischen Umsetzung schwieriger als dies bei „klassischen“ randomisierten Studien der Fall ist. Die Geheimhaltung der Zuordnung der Cluster (meist waren dies Arztpraxen) in Interventions- oder Kontrollgruppe bereitete hierbei in aller Regel keine Probleme, vielmehr war dies bei der nachfolgenden Rekrutierung der Patienten der

---

<sup>1</sup> Bei den 17 Studienprotokollen war – da diese lediglich von geplanten Studien berichteten – eine Beurteilung der Verzerrungsrisiken nicht möglich.

Fall. War nämlich den Studienpraxen zum Zeitpunkt der Patientenrekrutierung ihre Gruppenzuordnung bereits bekannt, so musste in Bezug auf das „allocation concealment“ ein hohes Verzerrungsrisiko angenommen werden. Dieses Problem hätte sich nur lösen lassen, indem man mit der Bekanntgabe der Gruppenzuteilung der Cluster (Studienpraxen) bis *nach* dem Abschluss der Patientenrekrutierung gewartet hätte. Häufig waren sich die Autoren dieser Problematik sogar bewusst, nahmen aber das höhere Verzerrungsrisiko zugunsten eines einfacheren Rekrutierungsprozesses in Kauf. Aller Wahrscheinlichkeit nach kam es dadurch bei manchen Studien zu einer Stichprobenverzerrung (engl. selection bias), welche sich z. B. durch eine (zahlenmäßig) ungleiche Verteilung der Patienten/Probanden in Interventions- und Kontrollgruppe oder relevante Unterschiede zwischen den Gruppen hinsichtlich der Patientencharakteristika äußern kann.

Da die clusterrandomisierten Studien häufig „prozessorientierte“ Interventionen (3.2.4.1) untersuchten, waren bei *keiner* von ihnen Patienten/Probanden bzw. Studienorganisatoren verblindet (engl. blinding of participants and personnel). Deshalb musste (mit einer Ausnahme) für alle clusterrandomisierten Studien von einem hohen Risiko der Ungleichbehandlung der Gruppen (engl. performance bias) ausgegangen werden.

Das Risiko der „Detektionsverzerrung“ (engl. detection bias) war trotz meist unverblindeter Zielgrößenerhebung (engl. blinding of outcome assessment) bei den eingeschlossenen clusterrandomisierten Studien verhältnismäßig ausgeglichen (7 Studien mit hohem, 3 mit unklarem und 7 mit niedrigem Risiko), da nicht selten objektive Zielkriterien verwendet wurden, für die auch bei mangelnder Verblindung nur geringe Verzerrungsrisiken bestehen.

Bei 5 Studien musste aufgrund einer hohen Anzahl vorzeitiger Studienabbrüche (engl. incomplete outcome data) ein hohes Verzerrungsrisiko (engl. attrition bias) angenommen werden, 8 Studien wiesen ein unklares und 4 ein niedriges Verzerrungsrisiko auf.

Das Verzerrungsrisiko aufgrund selektiver Berichterstattung (engl. selective reporting bzw. reporting bias) wurde für alle clusterrandomisierten Studien als niedrig eingestuft.

Zusammenfassend kann gesagt werden, dass jede der eingeschlossenen clusterrandomisierten Studien in mindestens einer Kategorie (also in mindestens einem der 6 Items, die in 2.6.1 bis 2.6.6 und in Tabelle 27 beschrieben sind) ein hohes Risiko vorwies.

#### **3.3.2 Verzerrungsrisiken der „klassischen“ randomisierten Studien**

Im Allgemeinen waren die methodischen Herausforderungen bei den 16 „klassischen“ randomisierten Studien geringer als bei den 17 clusterrandomisierten Studien.

Für keine der 16 „klassischen“ RCTs musste das Verzerrungsrisiko bei der Erstellung der Randomisierungssequenz (engl. random sequence generation) als hoch eingestuft werden (unklar in 5, niedrig in 11 Fällen). Ähnlich war die Risikobewertung bei der „Geheimhaltung der Gruppenzuordnung“ (engl. allocation concealment): Bei keiner der 16 „klassischen“ RCTs wurde von einem hohen Verzerrungsrisiko ausgegangen (ebenfalls unklar in 5 und niedrig in 11 Fällen). Eine Stichprobenverzerrung (engl. selection bias) kann zwar nicht ausgeschlossen werden, ist aber für die 16 „klassischen“ RCTs eher unwahrscheinlich.

Dagegen musste das Verzerrungsrisiko in Bezug auf eine mögliche ungleiche Behandlung/Versorgung innerhalb der beiden Gruppen (engl. performance bias) aufgrund einer fehlenden Verblindung von Patienten/Probanden und Studienpersonal (engl. blinding of participants and personnel) in 11 Fällen als „hoch“ eingestuft werden (von den verbleibenden Studien wurden 1 mit „unklar“ und 4 mit „niedrig“ bewertet).

Die erwähnte mangelnde Verblindung (von Studienteilnehmern und -personal) lag häufig zusammen mit subjektiven Zielkriterien vor, welche durch ebendiese unverblindeten Personen bewertet werden sollten. Somit war auch die Zielgrößenerhebung (engl. outcome assessment) unverblindet, sodass das Risiko der „Detektionsverzerrung“ (engl. detection bias) hoch war. Primär aus diesem Grund mussten 7 Studien bzgl. der Verblindung der „Outcomemessung“ (engl. blinding of outcome assessment) der hohen Risikokategorie zugeordnet werden (1 Studie „unklar“, 8 „niedrig“).

Die Höhe des Datenverlustes (engl. incomplete outcome data), welcher ebenfalls ein Verzerrungsrisiko darstellen kann (engl. attrition bias), war bei den untersuchten Studien variabel (hohes Verzerrungsrisiko in 3, unklares in 4 und niedriges in 9 Fällen). Grund hierfür war die Heterogenität der Studien: So gab es z. B. Kurzzeitstudien ohne jegliche Studienabbrecher, während bei anderen Studien die Abbrecherquote bei über 50 % lag. Außerdem hatte die Art der statistischen Auswertung (z. B. ob eine ITT-Analyse durchgeführt wurde oder nicht) einen Einfluss auf die Höhe dieses Verzerrungsrisikos.

In zwei Studien [Peters-Klimm et al., 2010b, Voigt et al., 2011] fanden sich Hinweise auf eine selektive Berichterstattung (engl. selective reporting), sodass das Verzerrungsrisiko (engl. reporting bias) als „hoch“ bewertet werden musste. Gründe waren eine Hervorhebung positiver Studienergebnisse bzw. eine inadäquate statistische Analyse, welche die Daten in ein falsches Licht rückte. Das Verzerrungsrisiko zweier weiterer Studien [Gastpar et al., 2003, Schencking et al., 2013] wurde in Bezug auf eine selektive Berichterstattung als „unklar“ eingestuft.

### 3 Ergebnisse

Tabelle 27: Verzerrungsrisiken der untersuchten Studien

Erstautor	Jahr	„random sequence generation“	„allocation concealment“	„blinding of participants and personnel“	„blinding of outcome assessment“	„incomplete outcome data“	„selective reporting“
<b>clusterrandomisierte Studien</b>							
Altiner	2007	niedrig	hoch	hoch	niedrig	hoch	niedrig
Becker	2008	niedrig	hoch	hoch	hoch	hoch	niedrig
Erler	2012	niedrig	niedrig	hoch	niedrig	niedrig	niedrig
Freiberger	2013	niedrig	hoch	hoch	hoch	unklar	niedrig
Gensichen	2009	niedrig	hoch	hoch	hoch	niedrig	niedrig
Junius-Walker	2012	unklar	hoch	hoch	hoch	unklar	niedrig
Kaufmann-Kolle	2011	unklar	hoch	hoch	unklar	hoch	niedrig
Krones	2008	unklar	hoch	hoch	hoch	unklar	niedrig
Peters-Klimm	2009	niedrig	niedrig	hoch	niedrig	unklar	niedrig
Mehring	2013	niedrig	hoch	hoch	niedrig	unklar	niedrig
Mehring	2014	niedrig	hoch	hoch	niedrig	unklar	niedrig
Rosemann	2007	unklar	niedrig	hoch	hoch	unklar	niedrig
Szecsényi	2011	unklar	unklar	hoch	niedrig	niedrig	niedrig
Tinsel	2013	niedrig	niedrig	unklar	unklar	hoch	niedrig
Vollmar	2007	unklar	hoch	hoch	niedrig	niedrig	niedrig
Vollmar	2010	unklar	niedrig	hoch	unklar	hoch	niedrig
Vormfelde	2014	niedrig	unklar	hoch	hoch	unklar	niedrig
<b>randomisierte Studien zu definierten Behandlungen</b>							
Bleidorn	2010	niedrig	niedrig	niedrig	niedrig	niedrig	niedrig
Bücker	2010	niedrig	niedrig	hoch	hoch	hoch	niedrig
Du Moulin	2009	niedrig	niedrig	hoch	niedrig	unklar	niedrig
Frese	2012	unklar	unklar	niedrig	niedrig	hoch	niedrig
Gastpar	2003	niedrig	niedrig	niedrig	niedrig	niedrig	unklar
Hensler	2009	unklar	niedrig	niedrig	niedrig	unklar	niedrig
Jobst	2005	niedrig	niedrig	hoch	hoch	unklar	niedrig
Klein	2013	niedrig	niedrig	hoch	niedrig	niedrig	niedrig
Peters-Klimm	2010	niedrig	niedrig	hoch	hoch	unklar	hoch
Schencking	2013	niedrig	niedrig	hoch	hoch	niedrig	unklar
Voigt	2011	unklar	unklar	hoch	hoch	niedrig	hoch
<b>sonstige randomisierte Studien</b>							
Bergold	2013	niedrig	niedrig	hoch	hoch	niedrig	niedrig
Blank	2013	niedrig	niedrig	hoch	niedrig	hoch	niedrig
Butzlaff	2004	niedrig	unklar	hoch	hoch	niedrig	niedrig
Hoffmann	2014	unklar	unklar	hoch	niedrig	niedrig	niedrig
Müller-Bühl	2011	unklar	unklar	unklar	unklar	niedrig	niedrig

Für Zuordnung von Studien zu Publikationen siehe „Übersicht aller Studienreferenzen“ im Literaturverzeichnis.

### 3.4 Zusammenfassung und Beurteilung der Studienergebnisse

Dieses Kapitel bezieht sich nur auf die 33 bereits abgeschlossenen Studien, da für die 17 Studienprotokolle noch keine abschließenden Ergebnisse vorlagen.

#### 3.4.1 „Vote count“

Für die 17 clusterrandomisierten Studien erfolgte hinsichtlich der „author’s conclusion“ in 5 Fällen eine Zuordnung zur Kategorie „+“, in 3 zur Kategorie „(+“ und in 9 zur Kategorie „0“ (2.7.1, Tabelle 28, Tabelle 30). Die „reviewer’s conclusion“ teilte 4 Studien der Kategorie „+“, 3 der Kategorie „(+“ und 10 der Kategorie „0“ zu. Somit bestanden 2 Abweichungen von „author’s“ und „reviewer’s conclusion“, wobei die „reviewer’s conclusion“ in diesen Fällen jeweils eine Bewertungskategorie unter der „author’s conclusion“ angesiedelt war.

Tabelle 28: „Vote count“-Zusammenfassung für clusterrandomisierte Studien (n = 17)

Bewertungskategorie	„author’s conclusion“	„reviewer’s conclusion“
+	5 Studien	4 Studien
(+)	3 Studien	3 Studien
0	9 Studien	10 Studien
(-)	-	-
-	-	-

Bei der „author’s conclusion“ in Bezug auf die 16 „klassischen“ RCTs wurden 4 Studien der Kategorie „+“ sowie jeweils 6 den Kategorien „(+“ und „0“ zugeordnet. Den Kategorien „+“, „(+“ und „0“ wurden bei der „reviewer’s conclusion“ 3, 5 und 8 Studien zugeteilt. Es lagen also 3 Abweichungen von „author’s“ und „reviewer’s conclusion“ (Tabelle 29 und Tabelle 30) vor. Auch in diesen drei Fällen wurden die jeweiligen Studien durch den „reviewer“ eine Bewertungskategorie niedriger eingestuft.

Tabelle 29: „Vote count“-Zusammenfassung für „klassische“ randomisierte Studien (n = 16)

Bewertungskategorie	„author’s conclusion“	„reviewer’s conclusion“
+	4 Studien	3 Studien
(+)	6 Studien	5 Studien
0	6 Studien	8 Studien
(-)	-	-
-	-	-

### 3 Ergebnisse

Tabelle 30: „Vote count“-Bewertungen der untersuchten Studien

Erstautor	Jahr	„author’s conclusion“	„reviewer’s conclusion“
<b>clusterrandomisierte Studien</b>			
Altiner	2007	+	+
Becker	2008	(+)	0
Erler	2012	(+)	(+)
Freiberger	2013	+	(+)
Gensichen	2009	(+)	(+)
Junius-Walker	2012	0	0
Kaufmann-Kolle	2011	0	0
Krones	2008	+	+
Peters-Klimm	2009	0	0
Mehring	2013	+	+
Mehring	2014	0	0
Rosemann	2007	0	0
Szecsényi	2011	0	0
Tinsel	2013	0	0
Vollmar	2007	0	0
Vollmar	2010	0	0
Vormfelde	2014	+	+
<b>randomisierte Studien zu definierten Behandlungen</b>			
Bleidorn	2010	0	0
Bücker	2010	(+)	0
Du Moulin	2009	+	+
Frese	2012	+	(+)
Gastpar	2003	(+)	(+)
Hensler	2009	0	0
Jobst	2005	0	0
Klein	2013	0	0
Peters-Klimm	2010	(+)	(+)
Schencking	2013	(+)	0
Voigt	2011	(+)	(+)
<b>sonstige randomisierte Studien</b>			
Bergold	2013	+	+
Blank	2013	+	+
Butzlaff	2004	0	0
Hoffmann	2014	0	0
Müller-Bühl	2011	(+)	(+)

Für Zuordnung von Studien zu Publikationen siehe „Übersicht aller Studienreferenzen“ im Literaturverzeichnis.

### 3.4.2 Metaanalyse

Für 15 der 17 clusterrandomisierten Studien konnten standardisierte mittlere Differenzen (engl. standardised mean differences, SMDs) berechnet werden (Tabelle 31). Da sich unter diesen Studien 3 dreiarmlige Studien befanden, wurden für diese jeweils zwei Berechnungen durchgeführt (jeweils „Interventionsgruppe 1 im Vergleich zur Kontrollgruppe“ und „Interventionsgruppe 2 im Vergleich zur Kontrollgruppe“), sodass insgesamt 18 Gruppenvergleiche erfolgten. Es zeigte sich in 2 Fällen (mit SMDs von 0,60 und 0,63) ein „moderater“ Effekt [Mehring et al., 2013, Vormfelde et al., 2014]. In den verbleibenden Fällen waren die Effekte (mit SMDs von -0,17 bis 0,30) „klein“. In 8 Fällen war die Null nicht im Konfidenzintervall enthalten, d. h. in diesen Fällen war der Unterschied zwischen Interventions- und Vergleichsgruppe statistisch signifikant ( $p\text{-Wert} \leq 0,05$ ).

Eine entsprechende Berechnung konnte zudem für 15 der 16 „klassischen“ randomisierten Studien durchgeführt werden. Da sich auch hierunter eine dreiarmlige Studie befand, konnten insgesamt 16 Gruppenvergleiche angestellt werden (Tabelle 31). Hierunter fand sich in 3 Fällen ein „großer“ Effekt [Bergold et al., 2013, Blank et al., 2013, Bleidorn et al., 2010], in den verbleibenden Fällen lagen die berechneten Werte der SMD bei  $< 0,3$ , was einem „kleinen“ Effekt entspricht. Nur in 4 Fällen (von 16) kam es mit einem  $p\text{-Wert}$  von  $\leq 0,05$  zu einem statistisch signifikanten Ergebnis.

### 3 Ergebnisse

Tabelle 31: Berechnete Effektgrößen der untersuchten Studien

Erstautor	Jahr	für Berechnung verwendetes Zielkriterium	SMD	LL	UL
<b>clusterrandomisierte Studien</b>					
Altiner	2007	Häufigkeit antibiotischer Verschreibungen	0,30	0,14	0,46
Becker	2008	Funktionskapazität: Gruppe 1 vs. Kontrolle	0,15	0,01	0,29
		Funktionskapazität: Gruppe 2 vs. Kontrolle	0,11	-0,03	0,25
Erlor	2012	Anzahl an zu hohen Dosierungen	0,23	-0,11	0,56
Freiberger	2013	Sturzrisiko/Mobilität	0,27	0,05	0,49
Gensichen	2009	Depressions-Scores	0,22	0,05	0,38
Krones	2008	Patientenbeteiligung und -zufriedenheit	0,23	0,10	0,35
Peters-Klimm	2009	Lebensqualität	-0,17	-0,50	0,16
Mehring	2013	Gewichtsreduktion	0,60	0,27	0,92
Mehring	2014	Raucherstatus	0,08	-0,61	0,77
Rosemann	2007	Lebensqualität: Gruppe 1 vs. Kontrolle	0,08	-0,09	0,25
		Lebensqualität: Gruppe 2 vs. Kontrolle	0,17	0,00	0,34
Szecsényi	2011	Anzahl erreichter Therapieziele	0,01	-0,04	0,07
Tinsel	2013	Patientenbeteiligung (SDM-Score)	0,07	-0,06	0,20
Vollmar	2007	Zeit bis Eintritt Pflegeheim/Tod (G1 vs. K)	0,08	-0,24	0,41
		Zeit bis Eintritt Pflegeheim/Tod (G2 vs. K)	-0,07	-0,38	0,25
Vollmar	2010	Wissenszuwachs	0,02	-0,28	0,33
Vormfelde	2014	Wissensstand	0,63	0,41	0,85
<b>randomisierte Studien zu definierten Behandlungen</b>					
Bleidorn	2010	Symptome	1,03	0,10	1,97
Bücker	2010	Funktionskapazität	0,28	-0,10	0,66
Du Moulin	2009	6-Minuten-Gehtest	0,15	-0,37	0,68
Frese	2012	Tod	0,14	0,05	0,22
Gastpar	2003	Scores für Angststörungen	0,15	-0,18	0,48
Hensler	2009	Krankheitsdauer	0,05	-0,32	0,41
Jobst	2005	Krankheitstage	0,02	-0,42	0,46
Klein	2013	Bewegungsrestriktion im Nacken	0,24	-0,27	0,74
Peters-Klimm	2010	Lebensqualität	-0,04	-0,38	0,31
Schencking	2013	Schmerz (Gruppe 1 vs. Kontrolle)	0,20	-0,68	1,08
		Schmerz (Gruppe 2 vs. Kontrolle)	0,10	-0,78	0,97
<b>sonstige randomisierte Studien</b>					
Bergold	2013	Wissenszuwachs	0,76	0,38	1,14
Blank	2013	Fähigkeiten der klinischen Untersuchung	1,16	0,58	1,74
Butzlaff	2004	Wissenszuwachs	0,11	-0,43	0,65
Hoffmann	2014	Fehlermanagement	-0,06	-0,57	0,45
Müller-Bühl	2011	Anzahl unbeantworteter Fragen	0,25	-0,02	0,52

Für Zuordnung von Studien zu Publikationen siehe „Übersicht aller Studienreferenzen“ im Literaturverzeichnis.

Abkürzungen (weitere Abkürzungen sowie Details zu den u. g. Abkürzungen siehe Abkürzungsverzeichnis): G1/G2 = Gruppe 1/2; K = Kontrolle; LL/UL = lower/upper limit of 95 % CI; SMD = standardized mean difference

## 4 Diskussion

### 4.1 Zusammenfassung der Hauptergebnisse

Innerhalb der untersuchten 15 Jahre wurden durch deutsche Universitätsabteilungen für Allgemeinmedizin 50 RCTs durchgeführt, wobei die Zahl der durchgeführten RCTs über den genannten Zeitraum stark zugenommen hat. So wurde keine der 79 eingeschlossenen Publikationen (zu den 50 untersuchten RCTs) in dem ersten Dreijahresintervall (2000–2002) des Untersuchungszeitraumes 2000–2014 veröffentlicht, während 35, also 44 %, auf das letzte Dreijahresintervall (2012–2014) entfielen (Abbildung 5). Von den 79 eingeschlossenen Publikationen waren 69 (87 %) in englischer Sprache verfasst und 70 (89 %) wiesen einen allgemeinmedizinischen Erstautor auf. 15 (42 %) der heute 36 allgemeinmedizinischen Universitätsabteilungen waren an mindestens einer der 50 eingeschlossenen RCTs beteiligt. Allerdings wurde der Großteil der Studien (39 von 50, also 78 %) durch eine verhältnismäßig kleine Gruppe von Instituten (6 von 36, also 17 %) durchgeführt. Bei fast der Hälfte der eingeschlossenen Studien (24 von 50) war ein Bundesministerium (BMBF oder BMG) der alleinige Geldgeber. Diejenigen RCTs, welche einen direkten Krankheitsbezug vorwiesen (34 von 50, also 68 %), wurden anhand des ICPC-2 Klassifikationssystems kategorisiert, wobei die Überkapitel „Psychologisch“, „Kreislauf“, „Bewegungsapparat“ und „Atmungsorgane“ am häufigsten vertreten waren. Unter den untersuchten RCTs stellten „prozessorientierte“ Studien (zum Teil sehr groß, meist clusterrandomisiert), deren Interventionen z. B. über Trainings- und „Case-management“-Programme auf die Optimierung von Vorgehensweisen in der Praxis oder durch den Patienten abzielten, einen Schwerpunkt dar – Medikamentenstudien dagegen die Ausnahme. 33 der 50 untersuchten RCTs waren abgeschlossene Studien, die verbleibenden 17 Studien befanden sich in Planung oder Durchführung (Studienprotokolle ohne Ergebnisse). Ein relevanter Teil der Studien wies hohe Verzerrungsrisiken (engl. bias) auf, was bei vielen Studien einer in der hausärztlichen Praxis nur schlecht durchführbaren Verblindung geschuldet war; bei clusterrandomisierten Studien war dies häufig auf eine mangelnde Geheimhaltung der Gruppenzuordnung (engl. allocation concealment) zurückzuführen. Bei den untersuchten Interventionen waren die Größen der Effekte im Vergleich zu den Kontrollen meist klein. Die 17 analysierten Studienprotokolle legen nahe, dass Vorerfahrungen mit RCTs, zunehmende methodische Expertise und Vernetzung zu einer weiteren qualitativen Verbesserung von zukünftigen allgemeinmedizinischen RCTs in Deutschland führen dürften.

## **4.2 Methodenkritik**

### **4.2.1 Literatursuche**

Eine Stärke der vorliegenden Arbeit stellt die Literatursuche dar, welche aufgrund der Verwendung mehrerer Suchstrategien (Datenbanksuche Scopus, Durchsicht der Publikationslisten aller deutschen Universitätsabteilungen für Allgemeinmedizin sowie „reference tracking“) die relevanten Publikationen weitestgehend vollständig erfasst haben sollte. Neben der Datenbanksuche erwies sich das „reference tracking“ als sehr effektiv, sodass über die zu einem späteren Zeitpunkt durchgeführte Durchsicht der Publikationslisten nur eine einzige zusätzliche Publikation gefunden werden konnte. Die zusätzliche Publikation [Frese et al., 2012] war dem Suchalgorithmus entgangen, weil dieser nur „Allgemeinmedizin“ bzw. „general practice“, nicht jedoch „primary care“ beinhaltete. Anhand dieses Beispiels wird auch das Hauptproblem bei der Erstellung eines Suchalgorithmus deutlich: Je vollständiger die Ergebnisse einer Suche ausfallen sollen, desto mehr irrelevante Treffer beinhaltet sie (welche nachträglich wieder aussortiert werden müssen). Dessen ungeachtet kann eine alleinige Datenbanksuche – auch in Hinblick auf zukünftige Arbeiten – nicht empfohlen werden. Grund hierfür sind häufig vorkommende unvollständige, fehlerhafte oder stark abgekürzte Affiliationsangaben, die das Auffinden einer Referenz durch den Suchalgorithmus unmöglich machen können. Eine auf mehreren Säulen fußende Literatursuche ist deshalb empfehlenswert, um möglichst vollständige Ergebnisse zu erzielen.

Die durchgeführte Scopussuche suchte nur nach Artikeln in deutscher und englischer Sprache. Artikel in anderen Sprachen wären somit nicht identifiziert worden. Die Ergebnisse von Grossmann [2014], die bereits in der Einleitung der vorliegenden Arbeit genannt wurden, deuten jedoch auf eine nur kleine Anzahl an relevanten Artikeln in anderen Sprachen als Deutsch und Englisch hin. So befanden sich unter den 794 bei Grossmann [2014] eingeschlossenen Artikeln genau zwei in französischer Sprache, während alle anderen in Deutsch oder Englisch verfasst waren. Zudem wären entsprechende Artikel in anderen Sprachen als Deutsch oder Englisch vermutlich über die beiden anderen Säulen der Literatursuche (Durchsicht Publikationslisten sowie „reference tracking“) identifiziert worden.

### **4.2.2 Erste Arbeit dieser Art – Stärke oder Schwäche?**

Dass es sich bei der vorliegenden Arbeit um die erste ihrer Art handelt kann als Stärke und Schwäche zugleich angesehen werden.

Vor der Erstellung dieser Übersichtsarbeit konnten die durch die universitäre deutsche Allgemeinmedizin durchgeführten RCTs nicht systematisch beurteilt und miteinander verglichen werden. Nun jedoch ist eine überblickende Zusammenschau möglich, die durchaus auch für zukünftige RCTs der deutschen Allgemeinmedizin von praktischer Bedeutung sein kann. So erlaubt der vorliegende Review einen kompakten Zugriff auf die aus vergangenen RCTs gezogenen Lehren, z. B. hinsichtlich des Rekrutierungsprozesses von Ärzten und Patienten, der benötigten Teilnehmerzahlen, drop-out-Raten, Drittmittelgeber, häufig begangener Fehler (die es zukünftig zu vermeiden gilt) etc. Auch in der akademischen Allgemeinmedizin unterrepräsentierte Forschungsfelder lassen sich anhand der vorliegenden Arbeit identifizieren, sodass diesen zukünftig besondere Aufmerksamkeit geschenkt werden kann.

Andererseits kann die Tatsache, dass diese Arbeit die erste ihrer Art ist, auch als Schwäche ausgelegt werden. So konnte nur in sehr eingeschränktem Maße auf Erfahrungswerte aus vorangegangenen Studien zurückgegriffen werden. Dies hatte zur Folge, dass im Verlauf der Arbeit aufkommende Grenzfälle, die bei der initialen Planung nicht vorhersehbar gewesen waren, manchmal eine Anpassung der methodischen Vorgehensweise nötig machten. Solche nachträglichen Änderungen (z. B. des Extraktionsformulars) inmitten des Extraktionsprozesses erzeugen unweigerlich eine gewisse Unschärfe.

### **4.2.3 Einschlusskriterien**

Die Analyse, von der die vorliegende Arbeit berichtet, hat lediglich Publikationen berücksichtigt, für deren Erst- und/oder Letztautoren eine Affiliation zu einem deutschen universitären Institut für Allgemeinmedizin bestand. Waren die Erst- und/oder Letztautoren jedoch deutsche Allgemeinmediziner ohne eine solche Affiliation, so erfolgte ein Ausschluss der entsprechenden Studie. Dies sollte allerdings keinen großen Einfluss auf die gewonnenen Ergebnisse haben, da der Fokus der vorliegenden Arbeit ohnehin auf durch die universitäre deutsche Allgemeinmedizin durchgeführte randomisierte Studien lag. Zudem dürfte die Anzahl an RCTs, die durch deutsche Allgemeinmediziner *ohne* universitäre Anbindung durchgeführt und publiziert wurden, sehr gering sein – dies wird auch durch den Abgleich mit den durch Kortekaas et al. [2012] zusätzlich identifizierten Publikationen gestützt (4.3).

Wenn in Publikationen zu randomisierten Studien, an denen deutsche Universitätsinstitute für Allgemeinmedizin beteiligt waren, lediglich „Mittelautoren“ die o. g. Voraussetzungen bzgl. ihrer Affiliation erfüllten, so wurden diese Studien nicht in die vorliegende Arbeit aufgenommen. Man könnte demnach argumentieren, dass ein möglicherweise relevanter

Anteil an RCTs, die unter Mitwirkung der deutschen universitären Allgemeinmedizin durchgeführt worden sind, unberücksichtigt blieb. Diese Argumentation lässt sich jedoch unschwer dadurch entkräften, dass in der vorliegenden Arbeit ohnehin primär solche Studien untersucht werden sollten, die *federführend* von deutschen Universitätsinstituten für Allgemeinmedizin durchgeführt worden waren. Hatten nun aber bei einer Publikation zu einer RCT die „allgemeinmedizinischen“ Autoren innerhalb der Autorenschaft „nur“ eine Mittelstellung inne, so wurde dadurch ebendiese Federführung in Frage gestellt, denn Erst- und/oder Letztautoren gehören aller Regel nach der Institution an, die das entsprechende Forschungsprojekt primär leitet.

Dennoch wird hieraus deutlich, dass die Beurteilung, unter welchen Voraussetzungen eine Studie oder eine Publikation als „allgemeinmedizinisch“ bzw. „der universitären deutschen Allgemeinmedizin zugehörend“ bezeichnet werden kann, schwierig ist. Die in der vorliegenden Arbeit verwendete Definition für allgemeinmedizinische Studien (Erst- und/oder Letztautor einer Publikation, die über eine RCT berichtet, müssen eine Affiliation zu einem deutschen Universitätsinstitut für Allgemeinmedizin vorweisen) kann also durchaus kritisch hinterfragt werden. Nichts desto trotz stellt die gewählte Vorgehensweise sicherlich eine der besten und praktikabelsten Lösungen dar.

### **4.2.4 Hürden und Kritikpunkte der Studienanalyse**

#### **4.2.4.1 Randomisierungsarten**

Überraschend war, dass es bei einzelnen Studien sogar schwierig war, deren Randomisierungsart („klassisch“ oder cluster) festzustellen. Im Allgemeinen kann gesagt werden, dass bei „klassischen“ randomisierten Studien die Erhebung der Zielkriterien auf derselben Ebene stattfindet, auf der auch die Randomisierung vorgenommen wurde. Im Vergleich dazu erfolgt bei clusterrandomisierten Studien die Zielgrößenerhebung auf einer anderen (meist untergeordneten) Ebene als derjenigen, auf welcher die Randomisierung stattgefunden hat (z. B. Randomisierung von Praxen, Zielgrößenerhebung bei Patienten). Diese Unterscheidung gestaltete allerdings sich in manchen Fällen schwierig. Als Beispiel kann eine in der vorliegenden Arbeit untersuchte Studie dienen [Löffler et al., 2014a], die auf die Reduzierung des Antibiotikaeinsatzes in Zahnarztpraxen abzielte. Hierfür wurden die Zahnarztpraxen der Interventionsgruppe im Anschluss an die Randomisierung zum Thema „rationaler Antibiotikaeinsatz“ geschult, während die Praxen der Kontrollgruppe keine Intervention erhielten. Als zentrales Zielkriterium galt die Anzahl verschriebener Antibiotika in dem Zeitraum von 6 Monaten nach der Intervention. Die Erhebung dieses Zielkriteriums kann allerdings sowohl auf Ebene der Praxis (z. B. durch Auswertung der

Verschreibungsdaten über die Praxissoftware) als auch auf Ebene der Patienten (z. B. indem man die Rezepte aller Patienten auswertet, welche die Praxis nach dem Zahnarztbesuch wieder verlassen) erfolgen. Werden entsprechende Detailinformationen zur Datenerhebung in der Publikation jedoch nicht genannt, so kann die Einteilung einer Studie in „klassisch“- oder clusterrandomisiert Schwierigkeiten bereiten.

#### **4.2.4.2 Heterogenität der Studien**

Das zentrale Problem bestand jedoch in der ausgeprägten Heterogenität der eingeschlossenen RCTs (Beispiel 7). Um trotzdem mittels Kategorisierungen, Auswertungen, Zusammenfassungen etc. einen Vergleich der Studien zu ermöglichen, mussten teilweise Kompromisse eingegangen werden, sodass nicht alle der extrahierten Detailinformationen in der Analyse berücksichtigt werden konnten. Besonders bei der Qualitätsbeurteilung der Studien (engl. Risk of bias assessment) stellte die genannte Heterogenität der Studien eine Herausforderung dar. Dies wird bei der Betrachtung der Extreme einiger der erhobenen Daten deutlich. Beispielsweise reichte die Studiendauer der untersuchten RCTs (also die Zeit, die vom Studieneinschluss eines Probanden bis zu dessen „Outcomemessung“ verging) von nur wenigen Stunden [Klein et al., 2013] bis hin zu vier Jahren [Vollmar et al., 2007a], während sich die Teilnehmerzahlen zwischen 20 [Du Moulin et al., 2009] und 46 000 [Löffler et al., 2014a] Probanden bewegten. Ob nun kleine Kurzzeitstudien, die nur wenige Stunden dauern und meistens keinen einzigen Studienabbrecher zu verzeichnen haben, (z. B. hinsichtlich der „drop-out-rate“) eine bessere Qualitätsbeurteilung erhalten sollten als groß angelegte, multizentrische RCTs, die mehrere Jahre dauern und deshalb naturgemäß Abbrecherquoten im zweistelligen Prozentbereich vorweisen, ist fraglich. Obwohl versucht wurde eine „faire“ und transparente Beurteilung nach möglichst objektiven Gesichtspunkten vorzunehmen, sind die Probleme, welche in der extremen Heterogenität der Studien begründet waren, als Schwachpunkt dieser Arbeit zu werten.

#### **4.2.4.3 Fehlende Zweitbeurteilung**

Auch wenn für immerhin ungefähr die Hälfte der Studien bei der Einschätzung der Verzerrungsrisiken eine zweite Beurteilung durch den Betreuer erfolgte, wurden große Teile der vorliegenden Arbeit, wie die Selektion der Studien, die Extraktion von Informationen, und die Zusammenfassung der Ergebnisse, durch den Autor alleine durchgeführt (der Betreuer wurde lediglich bei Unklarheiten hinzugezogen), was als weitere methodische Schwachstelle zu werten ist. Für eine unabhängige Bearbeitung aller Schritte der Analyse

nach dem Vier-Augen-Prinzip, die zweifelsohne sinnvoll gewesen wäre, fehlten allerdings die nötigen zeitlichen und personellen Ressourcen.

Deshalb (und aufgrund der anderen genannten methodischen Hürden) sollten die entsprechenden Beurteilungen lediglich als orientierend angesehen werden.

#### **4.2.4.4 Nachträglicher Studienausschluss**

Weiterhin könnte der nachträglich vorgenommene Ausschluss von zehn Studien bemängelt werden. Dieses Vorgehen wurde dennoch von Autor und Betreuer als sinnvoll erachtet, weil die entsprechenden Studien – trotz Erfüllung der vordefinierten Einschlusskriterien – aus verschiedenen Gründen (Tabelle 10) eindeutig nicht zu der zentralen Fragestellung dieser Arbeit passten.

### **4.3 Interpretation**

#### **4.3.1 RCTs der deutschen universitären Allgemeinmedizin – Entwicklung und Vergleich**

Wie zuvor erwähnt ist über den Untersuchungszeitraum von 2000–2014 ein deutlicher Anstieg an durch die deutsche universitäre Allgemeinmedizin durchgeführten und publizierten RCTs zu verzeichnen.

Der Anteil an RCTs unter allen in der akademischen Allgemeinmedizin in Deutschland durchgeführten Studien ist trotzdem gering. So wiesen von den 794 durch Grossmann [2014] identifizierten Arbeiten aus dem Zeitraum 2000–2010 nur 60 (7,6 %) ein randomisiertes Studiendesign auf<sup>1</sup>. In der Research Agenda des EGPRN (= European General Practice Research Network) wurde deshalb bereits 2009 eine vermehrte Realisierung von Interventionsstudien angeregt [Hummers-Pradier et al., 2009].

Zudem kann Deutschland – trotz des beschriebenen Anstiegs an RCTs – in Bezug auf randomisierte Studien nicht mit den „Spitzenländern“ der allgemeinmedizinischen Forschung, wie USA, Großbritannien, Skandinavien oder den Niederlanden, mithalten. Kortekaas et al. [2012] veröffentlichten eine Analyse des Outputs an allgemeinmedizinischen RCTs („RCTs with a general practice setting“) zwischen 1990 und 2010 auf internationaler Ebene. Von den 1935 eingeschlossenen Publikationen stammten 549 aus den USA, 511 aus Großbritannien, 201 aus Skandinavien, 194 aus den

---

<sup>1</sup> Da bei der vorliegenden Arbeit strengere Einschlusskriterien angewendet wurden, ist die Anzahl analysierter RCTs (trotz längerem Untersuchungszeitraum) geringer als bei Grossmann [2014].

Niederlanden und 480 aus anderen Ländern. Die Anzahl entsprechender Publikationen aus Deutschland war hierbei so gering, dass diese nicht einmal separat ausgewiesen wurden. Die Erstautorin stellte allerdings freundlicherweise die unveröffentlichte Liste der 52 Publikationen aus Deutschland zur Verfügung, von denen 38 in den Untersuchungszeitraum der vorliegenden Arbeit (2000–2010) fielen. Folglich mussten sich diese 38 durch Kortekaas et al. [2012] identifizierte Arbeiten auch in der vorliegenden Analyse wiederfinden und umgekehrt (d. h. die für den entsprechenden Zeitraum identifizierten Publikationen der vorliegenden Arbeit mussten auch in der Studie von Kortekaas et al. [2012] enthalten sein). Beides ist jedoch nur bedingt der Fall: So identifizierte die vorliegende Arbeit nur 10 (darunter 2 Studienprotokolle) der erwähnten 38 Publikationen, während sie weitere 10 Publikationen enthält, die in der Analyse von Kortekaas et al. [2012] fehlen. Der Hauptgrund für diese Diskrepanz besteht darin, dass bei 22 der durch Kortekaas et al. [2012] eingeschlossenen Arbeiten kein einziger Autor eine Affiliation zu einem deutschen Universitätsinstitut für Allgemeinmedizin vorwies (auch bestand für keinen dieser Autoren eine Affiliation zu einer Hausarztpraxis, sodass es sich hierbei vermutlich nicht einmal um Allgemeinärzte handelte); in 2 Publikationen hatten lediglich Mittelautoren eine entsprechende Affiliation. Die 4 verbleibenden Arbeiten waren entweder Zusatzpublikationen zu bereits eingeschlossenen Studien, Duplikate oder fälschlicherweise eingeschlossene Studien (die in Wirklichkeit gar keine RCTs waren, sondern über ein anderes Studiendesign verfügten). Somit liegt der Anteil der Überschneidungen bei lediglich 50 % (von den 20 in der vorliegenden Arbeit identifizierten Publikationen aus dem Zeitraum 2000–2010 waren 10 auch bei Kortekaas et al. [2012] enthalten). Dieser Vergleich belegt anschaulich, wie zwei Literatursuchen, die ein ähnliches Ziel verfolgen, aufgrund verschiedener methodischer Herangehensweisen, deutlich voneinander abweichende Ergebnisse erzielen können.

Auch eine weitere Analyse untermauert, dass Deutschland hinsichtlich des allgemeinmedizinischen RCT-Outputs nicht mit der kleinen Gruppe der führenden Länder der Allgemeinmedizinforschung mithalten kann. Diese Analyse verglich Quantität und Qualität des allgemeinmedizinischen Forschungsoutputs von Australien, Deutschland, Großbritannien, Kanada, den Niederlanden und der USA der Jahre 2001–2007 [Glanville et al., 2011]. Deutschland bildete hierbei nicht nur beim Publikationsoutput und der (anhand von Impact-Faktoren und Zitationsanalysen gemessenen) Forschungsqualität, sondern auch bei der „wissenschaftlichen Produktivität“ (die u. a. anhand von Forschungsausgaben

und der Anzahl wissenschaftlicher Mitarbeiter berechnet wurde) das Schlusslicht unter den sechs o. g. Ländern.<sup>1</sup>

### **4.3.2 Clusterrandomisierte Studien der universitären Allgemeinmedizinforschung**

Bei den 50 in der vorliegenden Arbeit untersuchten RCTs fiel auf, dass nur 20 (40 %) als „klassische“ randomisierte Studien konzipiert waren, während für 30 (60 %) Studien eine Clusterrandomisierung vorgesehen war. Ein so hoher Anteil an clusterrandomisierten Studien (unter allen RCTs) findet sich nur in sehr wenigen medizinischen Fachdisziplinen. So lieferte eine am 18.10.2016 durchgeführte orientierende Suche auf PubMed ([www.pubmed.com](http://www.pubmed.com)) für den Zeitraum („Date – Publication“) 01.01.2000–31.12.2014 insgesamt 261 765 Treffer, die als RCTs klassifiziert waren („Publication Type“ = „randomized controlled trial“). Wurden die genannten Suchkriterien um den Begriff „Cluster“ (in „All Fields“) erweitert, so wurden 4 285 Publikationen identifiziert. Dieser groben Abschätzung zufolge hätten clusterrandomisierte RCTs also nur einen Anteil von ca. 1–2 % an allen RCTs der medizinischen Forschung. Selbst unter Berücksichtigung der Ungenauigkeit dieser Schätzung erscheint ein Wert von unter 5 % realistisch. Der Anteil an clusterrandomisierten Studien (unter allen medizinischen RCTs) ist demnach in der deutschen Allgemeinmedizinforschung zweifelsohne überdurchschnittlich hoch.

Im Folgenden wird dargelegt, weshalb für randomisierte Studien der Allgemeinmedizin häufig eine Clusterrandomisierung gewählt wird. Dabei wird die Richtigkeit der Annahme, dass (randomisierte) Studien in ebendem „Setting“ stattfinden sollten, für das auch eine anschließende praktische Relevanz der Studienergebnisse vermutet wird, vorausgesetzt (andernfalls könnte die Übertragbarkeit der Ergebnisse in Frage gestellt werden; würde man beispielsweise allgemeinmedizinische Fragestellungen in Kliniken erforschen, so bliebe unklar, ob die gewonnenen Erkenntnisse tatsächlich auf den praktischen hausärztlichen Alltag anwendbar wären).

Daraus folgt, dass eine allgemeinmedizinische RCT in einer hausärztlichen Praxis durchgeführt werden sollte; in der Realität ist für entsprechende Studien meist sogar die Mitwirkung *vieler* Praxen notwendig (um die große Anzahl benötigter Patienten/Probanden rekrutieren zu können und um sicherzustellen, dass die Studienergebnisse nicht nur eine einzelne Studienpraxis, sondern die gesamte allgemeinmedizinische Versorgung in Deutschland abbilden). Ist dies – also die Beteiligung vieler Praxen an einer RCT –

---

<sup>1</sup> Deutschland schnitt beim Vergleich mit den stärksten Ländern der Allgemeinmedizinforschung schlecht ab. Im weltweiten Vergleich kann die Forschungsleistung der universitären Allgemeinmedizin in Deutschland natürlich trotzdem als gut bis sehr gut bezeichnet werden.

gegeben, so liegt die Durchführung einer Clusterrandomisierung auf der Hand, da sich Hausarztpraxen aufgrund ihrer Organisationsstruktur, Größe und örtlicher Trennung voneinander sehr gut als Cluster eignen und weil das Studiendesign der cRCT die weitere praktische Umsetzung der Studie erheblich vereinfacht. So können die beteiligten Praxen beispielsweise auf die oft aufwändige Randomisierung eines jeden einzelnen eingeschlossenen Patienten (mit ggf. Anruf im Studienzentrum etc.) verzichten und mit allen Studienpatienten gleichermaßen verfahren (was deutlich leichter ist, als für alle Patienten entsprechend deren Gruppenzuordnung zwei oder mehr Möglichkeiten in Diagnostik oder Therapie vorzuhalten). Auch für die zuvor angesprochenen (und in der allgemeinmedizinischen Forschung häufig anzutreffenden) „prozessorientierten“ Fragestellungen, die organisatorische Veränderungen oder edukative Interventionen untersuchen, eignen sich cRCTs besonders gut, da bei diesem Studiendesign das Risiko einer sog. Kontamination (d. h., dass Patienten der Kontrollgruppe fälschlicherweise doch einen Teil oder die gesamte Intervention des Interventionsarmes erhalten oder umgekehrt) deutlich niedriger ist als bei „klassischen“ RCTs [Chenot, 2009].

Es darf jedoch nicht außer Acht gelassen werden, dass clusterrandomisierte Studien auch einige Nachteile mit sich bringen. Anzunehmende Gemeinsamkeiten der Studienpatienten eines Clusters (hier: einer Praxis), die sich beispielsweise durch gleichartige Umwelteinflüsse, einen ähnlichen sozioökonomischen Status, vergleichbare Wohnverhältnisse oder berufliche Rahmenbedingungen äußern können (dies wird z. B. beim Vergleich des Patientenguts einer „Großstadtpraxis“ mit dem einer Landarztpraxis deutlich), bedingen eine sog. reduzierte Varianz (Streuung). Somit ist die Grundannahme der unabhängigen Patientenstichprobe nicht mehr gegeben. Dies hat zur Folge, dass bei einer cRCT (im Vergleich zur „klassischen“ RCT) eine viel größere Stichprobe (Patientenzahl) benötigt wird, um eine vergleichbare statistische Aussagekraft (Power) zu erzielen. Um den Besonderheiten einer Clusterrandomisierung Rechnung zu tragen sind weiterhin spezielle Analysen notwendig, wodurch sich die statistische Auswertung besonders aufwändig gestaltet und die Mitarbeit eines Statistikers häufig unabdingbar ist. Zudem bereitet eine Verblindung bei cRCTs mehr Probleme als bei „klassischen“ RCTs (eine adäquate Verblindung ist zwar theoretisch auch bei cRCTs möglich, die im Rahmen von cRCTs untersuchten Interventionen lassen aber häufig keine Verblindung zu). Ähnlich verhält es sich bei den Verzerrungsrisiken (engl. bias): Eine cRCT könnte zwar so gestaltet werden, dass die Verzerrungsrisiken gering ausfallen; in der Realität werden jedoch bei der methodischen Studiendurchführung (zugunsten eines vereinfachten Studienablaufs) oft Abstriche in Kauf genommen (z. B. beim „allocation concealment“, 3.3.1), sodass bei vielen cRCTs hohe Verzerrungsrisiken zu beobachten sind. [Chenot, 2009]

### 4.3.3 Inhaltliche Einordnung der analysierten Studien

Unter den Studien, denen eine medizinische Indikation zugeordnet werden konnte, stammte mehr als jede fünfte aus dem Bereich „Psychologisch“, knapp gefolgt von den Kategorien „Kreislauf“, „Bewegungsapparat“, und „Atmungsorgane“. Da mehr als ein Viertel aller Patienten, die einen Hausarzt aufsuchen, zum Zeitpunkt der Konsultation (bzw. während des vorangegangenen Monats) an einer definierten psychischen Erkrankung nach ICD-10 leiden [Maier et al., 1996] und laut DAK-Bericht 2009 [DAK, 2009] psychische und muskuloskelettale Erkrankungen neben Atemwegserkrankungen und Verletzungen die häufigsten Ursachen für Krankschreibungen (Arbeitsunfähigkeit) darstellen, scheinen die realen allgemeinmedizinischen Beratungsanlässe durch die o. g. RCT-Fragestellungen verhältnismäßig gut abgebildet zu werden.

Über die Hälfte der eingeschlossenen Studien (31 von 50, also 62 %) untersuchten sog. „prozessorientierte“ Interventionen, welche meist auf die Verbesserung von „Case-management“-Programmen, Abläufen innerhalb der Praxen o. ä. abzielten. Der Anteil solcher „prozessorientierter“ Studien war dabei unter den Studienprotokollen (15 von 17, also 88 %) deutlich höher als unter den bereits abgeschlossenen RCTs (16 von 33, also 48 %), was darauf hindeutet, dass Studien zu „prozessorientierten“ Interventionen auch zukünftig einen besonderen Schwerpunkt darstellen werden. Diese zukünftige Fokussierung auf komplexe, „prozessorientierte“ Interventionen (und weniger auf eng umschriebene spezifische Interventionen, wie dies beispielsweise bei Medikamentenstudien der Fall wäre) geht mit der Forschungsagenda der EGPRN einher, welche RCTs zu „größeren“ Themen wie Patientenpräferenzen, Multimorbidität, Lebensqualität sowie Sozial- und Umweltfaktoren fordert [Hummers-Pradier et al., 2009].

Es fällt zudem auf, dass für die genannten „prozessorientierten“ Studien überproportional häufig eine Clusterrandomisierung gewählt wurde. Im Gegensatz dazu zeigte sich bei den Studien zu „definierten Behandlungen“ deutlich häufiger das Studiendesign der „klassischen“ RCT.

Unter den Studien zu „definierten Behandlungen“ stach besonders heraus, dass sich fast die Hälfte der Studien (6 von 14) mit komplementärmedizinischen bzw. naturheilkundlichen Themen befasste. Es ist zwar bekannt, dass komplementär- und alternativmedizinische Heilverfahren in Deutschland (vor allem im ambulanten Bereich) aus verschiedenen Gründen sehr verbreitet sind [Joos et al., 2010, Linde et al., 2015], dennoch scheinen komplementärmedizinische Fragestellungen innerhalb der universitären allgemeinmedizinischen Forschung in Deutschland überrepräsentiert zu sein.

#### 4.3.4 Methodik der analysierten Studien

##### 4.3.4.1 Clusterrandomisierte Studien

Wie in 3.3.1 beschrieben, lag für jede der analysierten clusterrandomisierten Studien in mindestens einer der untersuchten Kategorien ein hohes Risiko einer Studienverzerrung vor; in den meisten Fällen bestanden jedoch auch noch in weiteren Risikokategorien „hohe“ oder zumindest „unklare“ Verzerrungsrisiken. In den Kategorien, für die bei cRCTs sehr häufig ungünstige Bewertungen vergeben werden mussten, handelte es sich um die Verblindung von Patienten/Probanden und Studienpersonal (engl. blinding of participants and personnel) sowie die Geheimhaltung der Gruppenzuordnung (engl. allocation concealment). Etwas seltener, aber immer noch in bedeutendem Maße, waren bei den cRCTs die Verblindung der Zielgrößenerhebung (engl. blinding of outcome assessment) sowie vorzeitige Studienabbrüche (engl. incomplete outcome data) mit hohen Verzerrungsrisiken belastet.

Die hohen Verzerrungsrisiken bei den eingeschlossenen cRCTs dürfen jedoch nicht grundsätzlich als Ausdruck einer schlechten Studienplanung oder -durchführung gewertet werden. Vielmehr sind sie als Schwachstellen von cRCTs anzusehen, die allgemein bekannt sind und bei vielen Studien – nicht nur bei allgemeinmedizinischen – Herausforderungen darstellen [Brierley et al., 2012, Díaz-Ordaz et al., 2014, Eldridge et al., 2009, Eldridge et al., 2004]. Die Verblindung von Patienten/Probanden und Studienpersonal (engl. blinding of participants and personnel), die beim „Risk of bias assessment“ der cRCTs das am ungünstigsten bewertete Item darstellte, kann hierfür als Beispiel dienen (mit nur einer Ausnahme musste hier für alle analysierten cRCTs von einem hohen Verzerrungsrisiko ausgegangen werden; Tabelle 27). Der Grund für die ausgeprägte Risikoanfälligkeit dieses Items liegt nämlich hauptsächlich darin, dass allgemeinmedizinische cRCTs oft „prozessorientierte“ Interventionen (3.2.4.1) – wie z. B. „Case-management“-Programme, Schulungen etc. – untersuchen (3.2.5 und Tabelle 24), bei denen sich Patienten/Probanden und Studienpersonal von Natur aus nicht (oder nur schlecht) verblenden lassen. Wird in einer solchen Studie nun zusätzlich noch ein subjektiver Zielparameter (z. B. Lebensqualität, Symptomschwere etc.) untersucht, der bei den unverblindeten Patienten/Probanden erhoben wird, so ist eine verblindete „Outcomemessung“ (engl. blinding of outcome assessment) ebenfalls nicht möglich. Für diese beiden Items („blinding of participants and personnel“ und „blinding of outcome assessment“) kann es also für bestimmte Fragestellungen – selbst bei bestmöglicher Studienplanung – *unmöglich* sein, ein niedriges Verzerrungsrisiko zu erzielen. Dennoch erscheint es auffällig, dass 90 % der in dieser Analyse untersuchten cRCTs keinerlei

Verblindung aufwiesen (3.2.5). Vermutlich ließe sich – bei angemessener Berücksichtigung dieser beiden Items („blinding of participants and personnel“ und „blinding of outcome assessment“) bereits in der Studienkonzeption – ein größerer Anteil an cRCTs (zumindest teilweise) verblinden, was mit einer Senkung der Verzerrungsrisiken einhergehen würde.

Im Gegensatz zu den beiden zuvor genannten Items („blinding of participants and personnel“ und „blinding of outcome assessment“), bei denen es – wie erwähnt – *unmöglich* sein kann die Verzerrungsrisiken klein zu halten, ist es in Bezug auf das „allocation concealment“ bei entsprechender Planung immer und bezüglich fehlender Daten („incomplete outcome data“) meistens *möglich* die Verzerrungsrisiken niedrig zu halten. Hinsichtlich des „allocation concealments“ müsste hierfür bei cRCTs mit der Bekanntgabe der Gruppenzuteilung der Cluster (hier: Praxen) bis *nach* dem Abschluss der Patientenrekrutierung gewartet werden, worauf allerdings aus Gründen der Praktikabilität häufig verzichtet wird (3.3.1). Einem hohen Anteil an fehlenden Daten könnte ebenfalls, beispielsweise durch eine engmaschige Nachverfolgung der Patienten/Probanden oder entsprechend hohe Anreizzahlungen, effektiv entgegengewirkt werden; gänzlich vollständige Datensätze werden aber vermutlich dennoch nur in seltenen Fällen erzielt werden können. Für die meisten randomisierten Studien gilt: Je länger die Studie dauert, desto größer ist die Studienabbrecherquote und entsprechend höher fällt der Anteil an fehlenden Daten (und das damit einhergehende Verzerrungsrisiko) aus. Dies ist insofern ungünstig, da gerade randomisierte Studien mit langen Nachbeobachtungszeiten von großer praktischer Relevanz und somit von besonderem Interesse sind. Deshalb sollten auch weitere Möglichkeiten zur Senkung des Verzerrungsrisikos durch fehlende Daten Anwendung finden. Ein Beispiel hierfür wäre ein statistisches Verfahren, die sog. ITT-Analyse, welche allerdings nur bei ca. der Hälfte der untersuchten cRCTs Anwendung fand (3.2.5).

### **4.3.4.2 „klassische“ randomisierte Studien**

Die methodischen Herausforderungen bei den „klassischen“ randomisierten Studien waren im Allgemeinen geringer als bei den clusterrandomisierten Studien (3.3.2). Kategorien mit ungünstigen Risikobewertungen waren hier – wie schon bei den clusterrandomisierten Studien (4.3.4.1) – die Verblindung von Patienten/Probanden und Studienpersonal (engl. blinding of participants and personnel) sowie die Verblindung der Zielgrößenerhebung (engl. blinding of outcome assessment), welche jedoch teilweise nicht umsetzbar sind. Die Ausführungen aus 4.3.4.1 gelten hierfür analog.

#### **4.3.4.3 Zukünftige RCTs der deutschen universitären Allgemeinmedizin**

Natürlich fanden sich im Rahmen der vorliegenden Arbeit – neben den in 4.3.4.1 und 4.3.4.2 beschriebenen, teilweise schlecht beeinflussbaren methodischen Hürden – auch Studien mit konzeptionellen Schwächen und Problemen bei der Studiendurchführung. Die untersuchten Studienprotokolle deuten jedoch darauf hin, dass die neueren (geplanten oder bereits angelaufenen) RCTs größer und professioneller werden. So findet sich neben steigenden Zahlen bei Autoren, beteiligten allgemeinmedizinischen Universitätsabteilungen und Probanden auch zunehmend bereits in der Phase der Studienplanung ein vorausschauender Umgang mit methodischen Herausforderungen (z. B. indem eine Clusterrandomisierung in Bezug auf Fallzahlschätzung und Wahl der Analysemethoden angemessen berücksichtigt wird). Langfristig sollte es demnach zu einer Qualitätssteigerung und einer damit einhergehenden Senkung von Verzerrungsrisiken kommen. Für diesen sich abzeichnenden Trend sind vermutlich auch im Rahmen früherer RCTs gesammelte Erfahrungen von Bedeutung.

#### **4.3.5 Effekte der untersuchten Interventionen**

Die im Rahmen der vorliegenden Arbeit ermittelten Effektgrößen zeigen, dass die Interventionseffekte der eingeschlossenen Studien häufig gering ausfielen. Um herauszufinden, ob bei internationalen allgemeinmedizinischen RCTs ebenfalls nur kleine Effekte erzielt werden konnten, sollte ein entsprechender Vergleich vorgenommen werden. Allerdings ließ sich keine systematische Übersichtsarbeit finden, welche die Effektgrößen von internationalen allgemeinmedizinischen randomisierten Studien themenübergreifend (also nicht nur in Bezug auf eine Untergruppe allgemeinmedizinischer RCTs, wie z. B. Studien zum Themenkomplex „depressive Störungen in der Allgemeinmedizin“) analysierte und demnach als Vergleichsgrundlage hätte dienen können. Die Interpretation der ermittelten niedrigen Effektgrößen beschränkt sich deshalb auf die folgende Einschätzung von Autor und Betreuer der vorliegenden Arbeit: Obwohl es grundsätzlich nicht verwunderlich ist, dass – bei den vielen im primärärztlichen Setting anzutreffenden Einflussfaktoren – im Rahmen von RCTs häufig nur geringe Interventionseffekte erzielt werden, sind die geringen Effektgrößen bei manchen der analysierten Studien vermutlich auch auf konzeptionelle Schwächen zurückzuführen.

### **4.4 Schlussfolgerung**

Nicht nur ist die Anzahl der durch die deutsche universitäre Allgemeinmedizin durchgeführten randomisierten Studien innerhalb der letzten 15 Jahren deutlich

angestiegen, sondern sie wurden auch (hinsichtlich Patientenzahlen und beteiligten Hausärzten) größer und, aufgrund methodischer Verbesserungen und damit einhergehender geringerer Verzerrungsrisiken, qualitativ hochwertiger. Dennoch kann Deutschland mit den führenden Ländern der Allgemeinmedizinforschung bei Weitem nicht mithalten. Um hier Anschluss zu finden, wären umfangreiche Maßnahmen nötig. Zunächst müsste es an allen medizinischen Fakultäten in Deutschland einen Lehrstuhl für Allgemeinmedizin und ein daran angeschlossenes, personell und materiell gut ausgestattetes Forschungsinstitut geben. Anschließend müssten die Institute den Aufbau von lokalen Forschungspraxis-Netzwerken vorantreiben, um nicht für jede einzelne randomisierte Studie in einem aufwändigen und kostenintensiven Rekrutierungsprozess neue Praxen für eine Studienbeteiligung gewinnen und schulen zu müssen. Für die Aufrechterhaltung des Netzwerks wären regelmäßige Forschungsprojekte unter Miteinbeziehung der Praxen notwendig.

Die Verwirklichung dieser Maßnahmen erfordert allerdings eine deutlich bessere finanzielle Ausstattung der deutschen universitären Allgemeinmedizin als dies bisher der Fall ist. Die Institute benötigen Gelder, um die Kosten für ein multiprofessionelles Forscherteam (z. B. mit Ärzten, Psychologen, Statistikern, Assistenzpersonal etc.), das zur Konzeption und Betreuung einer RCT benötigt wird, decken zu können. Weiterhin werden Gelder für die Etablierung und Aufrechterhaltung der Praxisnetze<sup>1</sup>, die geplanten Interventionen und die Anreizzahlungen benötigt.

Zusammenfassend kann gesagt werden, dass die langfristige Sicherstellung von Finanzmitteln in ausreichender Höhe als wichtigste Maßnahme zur nachhaltigen Stärkung von allgemeinmedizinischen Interventionsstudien in Deutschland anzusehen ist.

---

<sup>1</sup> Schätzungen, die auf Erfahrungen aus dem Ausland basieren, gehen von Kosten in Höhe von 300 000–500 000 € pro Jahr und lokalem Forschungsnetz aus. Eine bundesweite zentrale Koordinierungsstelle würde in etwa denselben finanziellen Rahmen benötigen. Zur Finanzierung der Infrastruktur von initial 8 Praxisnetzen in Deutschland wären somit jährlich ca. 4,5 Mio. € erforderlich. [Gágyor, 2014]

## 5 Zusammenfassung

**Hintergrund:** Klinische Forschung in der Allgemeinmedizin ist in Deutschland bis heute unterrepräsentiert. Um dies zu ändern wurde das DFG-Netzwerk „Klinische Studien in der Allgemeinmedizin“ ins Leben gerufen, welches die Durchführung des vorliegenden systematischen Reviews anregte. Ziel der Arbeit ist es, die durch deutsche Universitätsinstitute für Allgemeinmedizin durchgeführten randomisierten kontrollierten Studien (RCTs) der Jahre 2000–2014 hinsichtlich ihrer Fragestellungen, Methoden, Qualität und Ergebnisse zu analysieren. Die Ergebnisse können dem o. g. DFG-Netzwerk als Ausgangspunkt für künftige Verbesserungen dienen.

**Methoden:** In die Analyse wurden zwischen 2000 und 2014 veröffentlichte RCTs und publizierte Protokolle von RCTs eingeschlossen, bei denen der Erst- und/oder der Letztautor an eine Allgemeinmedizinabteilung einer deutschen Universität angebunden war(en). Die Identifikation von RCTs erfolgte über Suchen in der Literaturodatenbank Scopus, die Durchsicht verfügbarer Publikationslisten der Institute sowie die Nachverfolgung von Textreferenzen aus bereits eingeschlossenen Artikeln („reference tracking“). Die Beurteilung der Verzerrungsrisiken erfolgte anhand des Bewertungsinstruments der Cochrane Collaboration („The Cochrane Collaboration’s tool for assessing risk of bias“). Die in den untersuchten Studien erzielten Effektgrößen wurden anhand der berechneten standardisierten Mittelwertsdifferenzen (engl. standardised mean differences oder SMDs) quantifiziert.

**Ergebnisse:** Die 50 eingeschlossenen Studien (in 79 Publikationen) umfassten 33 abgeschlossene RCTs (in 62 Publikationen), zu denen bereits mindestens erste Ergebnisse verfügbar waren, sowie 17 Studien (in 17 publizierten Studienprotokollen), die sich noch im Stadium der Planung oder Durchführung befanden, sodass keine Ergebnisse vorlagen. Die Zahl der publizierten Studien und Studienprotokolle nahm zwischen den Jahren 2000 und 2014 stark zu. Von den 79 eingeschlossenen Publikationen wiesen 70 (89 %) einen allgemeinmedizinischen Erstautor auf, 69 (87 %) waren in englischer Sprache verfasst. Bei fast der Hälfte der eingeschlossenen Studien (24 von 50) war ein Bundesministerium (BMBF oder BMG) der alleinige Geldgeber. 15 (42 %) der heute 36 allgemeinmedizinischen Universitätsabteilungen waren an mindestens einer der 50 eingeschlossenen RCTs beteiligt.

Unter den 33 abgeschlossenen RCTs befanden sich 17 (zum Teil sehr große) clusterrandomisierte Studien (cRCTs). Diese untersuchten vornehmlich Interventionen, welche die Optimierung von Vorgehensweisen in der Praxis oder durch den Patienten (z. B. durch Trainings- oder „Case-management“-Programme) zum Ziel hatten (sog. „prozessorientierte“ Interventionen). Bei diesen 17 cRCTs lag die Anzahl an Clustern zwischen 6 und 606, die Teilnehmerzahlen reichten von 168 bis 7 807 Patienten/Probanden. Das Risiko von Verzerrungen (engl. bias) war – vorwiegend aufgrund der Gefahr einer Stichprobenverzerrung (engl. selection bias) infolge einer erst nach der Randomisierung der Cluster erfolgten Probandenrekrutierung – hoch. Die Interventionseffekte der cRCTs fielen im Vergleich zu den Kontrollen meist klein aus.

Bei den anderen 16 abgeschlossenen Studien handelte es sich um „klassische“ RCTs (d. h. die Randomisierung erfolgte auf derselben Ebene, auf der auch die Zielvariablen bestimmt wurden, vgl. 2.2.2) mit Teilnehmerzahlen zwischen 20 und 1 620 Patienten/Probanden. Sie untersuchten eine Vielzahl verschiedener Fragestellungen, wobei komplementär- und alternativmedizinische Themen auffällig häufig (6 von 16) vertreten waren. Die Verzerrungsrisiken dieser Studien waren sehr variabel; für eine ungünstige Bewertung war meist eine fehlende oder mangelhafte Verblindung verantwortlich. Bei Versorgungsforschungsstudien in hausärztlichen Praxen ist eine Verblindung allerdings häufig schlichtweg nicht umsetzbar und kann deshalb nicht zwingend als Ausdruck einer schlechten Studienplanung oder -durchführung gewertet werden. Die Interventionseffekte waren auch unter den „klassischen“ randomisierten Studien klein.

Bei den 17 Studien ohne Ergebnisse, für die lediglich Protokolle vorlagen, war sowohl der Anteil an (geplanten) Clusterrandomisierungen mit 13 von 17 Studien als auch die (geplanten) Probandenzahlen mit bis zu 46 000 Studienteilnehmern deutlich höher als bei den 33 bereits abgeschlossenen RCTs. Die Studienprotokolle deuten darauf hin, dass durch zunehmende methodische Expertise (z. B. angemessenere Berücksichtigung einer Clusterrandomisierung bei Studienplanung und -analyse) und Vorerfahrungen mit RCTs die Verzerrungsrisiken bei zukünftigen Studien weiter abnehmen werden. Da für diese Studien noch keine Ergebnisse vorlagen, konnten die Effektgrößen nicht beurteilt werden.

**Schlussfolgerung:** Trotz begrenzter Finanzmittel ist neben einem Anstieg der durch deutsche Universitätsabteilungen für Allgemeinmedizin durchgeführten RCTs auch eine Qualitätssteigerung derselben zu beobachten. Dennoch bedarf es verstärkter Anstrengungen und eines deutlich höheren finanziellen Mitteleinsatzes, um im internationalen Vergleich mithalten zu können.

## Abbildungsverzeichnis

Abbildung 1:	Selektion der Publikationen aus Vorgängerdissertationen (2000–2010) .....	15
Abbildung 2:	Selektion der Publikationen aus Datenbanksuche Scopus (2010–2014) .....	17
Abbildung 3:	Praktisches Vorgehen zur Identifizierung zusammengehörender Artikel.....	18
Abbildung 4:	Flowchart zur schematischen Darstellung des Selektionsprozesses....	38
Abbildung 5:	Gruppierung der 79 eingeschlossenen Publikationen nach dem Zeitpunkt ihrer Veröffentlichung (in 3-Jahres-Abschnitten).....	43
Abbildung 6:	Extraktionsformular, Seite 1 .....	xxxiv
Abbildung 7:	Extraktionsformular, Seite 2 .....	xxxv
Abbildung 8:	Extraktionsformular, Seite 3 .....	xxxvi
Abbildung 9:	Extraktionsformular, Seite 4 .....	xxxvii
Abbildung 10:	Liste für Schlagwortsuche .....	xxxviii
Abbildung 11:	Beurteilung der Verzerrungsrisiken anhand von ITT-Analyse und „fehlenden Daten“ .....	xxxix



## Tabellenverzeichnis

Tabelle 1:	Suchstrategie für Literatursuche Scopus .....	13
Tabelle 2:	„Random sequence generation“ – Beurteilung des Verzerrungsrisikos	25
Tabelle 3:	Zuteilung der Asthmapatienten in Interventions- bzw. Kontrollgruppe.	27
Tabelle 4:	„Allocation concealment“ – Beurteilung des Verzerrungsrisikos .....	28
Tabelle 5:	„Blinding of participants and personnel“ – Beurteilung des Verzerrungsrisikos .....	29
Tabelle 6:	„Blinding of outcome assessment“ – Beurteilung des Verzerrungsrisikos .....	31
Tabelle 7:	„Incomplete outcome data“ – Beurteilung des Verzerrungsrisikos.....	32
Tabelle 8:	„Selective reporting“ – Beurteilung des Verzerrungsrisikos .....	33
Tabelle 9:	Beurteilungsgrundlage für „vote count“ .....	34
Tabelle 10:	Ausgeschlossene Studien (n = 10) .....	41
Tabelle 11:	Veröffentlichungszeiträume, Publikationssprache und Autoren.....	44
Tabelle 12:	Internationale Autoren oder Mitautoren aus der Biometrie .....	44
Tabelle 13:	Beteiligung der einzelnen allgemeinmedizinischen Institute an den eingeschlossenen RCTs .....	46
Tabelle 14:	Finanzierung der eingeschlossenen Studien.....	48
Tabelle 15:	Einteilung der eingeschlossenen Studien nach Art der Intervention.....	50
Tabelle 16:	Einteilung der eingeschlossenen Studien nach Krankheitsbezug.....	51
Tabelle 17:	Einteilung der Studien mit direktem Krankheitsbezug (n = 34) nach ICPC-2 und ICD-10 .....	52
Tabelle 18:	Untersuchte Krankheiten im Freitext .....	53
Tabelle 19:	Abgeschlossene RCTs – clusterrandomisiert (n = 17).....	54
Tabelle 20:	Abgeschlossene RCTs – „klassisch“ randomisiert (n = 16).....	55
Tabelle 21:	Studienprotokolle ohne Ergebnisse (n = 17) .....	56
Tabelle 22:	Randomisierungsart der eingeschlossenen RCTs .....	57
Tabelle 23:	Studiencharakteristika, gruppiert nach der Art der Randomisierung ....	62
Tabelle 24:	Einteilung der eingeschlossenen Studien nach Interventions- und Randomisierungsart .....	63
Tabelle 25:	Anzahl Studienzentren und Durchführungsorte der eingeschlossenen Studien .....	64
Tabelle 26:	Anzahl der Patienten/Probanden zu Studienbeginn (t <sub>0</sub> ) und zum Zeitpunkt der ersten Datenerhebung (t <sub>1</sub> )* .....	67
Tabelle 27:	Verzerrungsrisiken der untersuchten Studien .....	72

Tabelle 28:	„Vote count“-Zusammenfassung für clusterrandomisierte Studien (n = 17) .....	73
Tabelle 29:	„Vote count“-Zusammenfassung für „klassische“ randomisierte Studien (n = 16) .....	73
Tabelle 30:	„Vote count“-Bewertungen der untersuchten Studien.....	74
Tabelle 31:	Berechnete Effektgrößen der untersuchten Studien .....	76
Tabelle 32:	Aufschlüsselung der 95 Publikationen nach ihrer Zugehörigkeit zu den 60 RCTs .....	xxxii
Tabelle 33:	Anzahl der beteiligten Praxen/Ärzte zu Studienbeginn ( $t_0$ ) und zum Zeitpunkt der ersten Datenerhebung ( $t_1$ )* .....	xli

## Abkürzungsverzeichnis

A.	lat. Arteria (deutsch: Arterie)
Abb.	Abbildung
abh.	abhängig
AM	Allgemeinmedizin
Anz.	Anzahl
AOK	Allgemeine Ortskrankenkasse
apl. Prof.	außerplanmäßiger Professor
ASI	engl. Anxiety Status Inventory (Instrument zur Bewertung von Angsterkrankungen)
BÄK	Bundesärztekammer
BAT	Bundesangestelltentarif
Blutungskompl.	Blutungskomplikation
BMBF	Bundesministerium für Bildung und Forschung
BMG	Bundesministerium für Gesundheit
BMI	engl. Body-Mass-Index (deutsch: Körpermasseindex)
BZ	Blutzucker
bzgl.	bezüglich
bzw.	beziehungsweise
c-o	engl. Cross-Over (spezielles Studiendesign)
CAI	engl. Colitis Activity Index (Aktivitätsindex für Colitis ulcerosa)
CDAI	engl. Crohn's Disease Activity Index (Aktivitätsindex für Morbus Crohn)
chron.	chronisch
CI	engl. confidence interval (deutsch: Konfidenzintervall)
CIMT	engl. Constraint Induced Movement Therapy (Therapieverfahren, das v. a. im Rahmen der Rehabilitation nach Schlaganfällen angewendet wird)
CME	engl. Continuing Medical Education (kontinuierliche berufsbegleitende Fortbildung für Ärzte)
CO	Kohlenmonoxid
CO <sub>2</sub>	Kohlenstoffdioxid
COPD	engl. Chronic Obstructive Pulmonary Disease (deutsch: Chronisch obstruktive Lungenerkrankung)

cRCT	cluster-RCT
d	Studiendauer
d. h.	das heißt
D. m.	Diabetes mellitus
DAK	DAK-Gesundheit (Versicherungsträger der deutschen gesetzlichen Krankenversicherung)
DEGAM	Deutsche Gesellschaft für Allgemeinmedizin
DFG	Deutsche Forschungsgemeinschaft
DMP	engl. Disease-management-program (organisiertes Behandlungsprogramm für chronisch kranke Menschen)
€	Euro
EGPRN	engl. European General Practice Research Network (deutsch: Europäisches Forschungsnetzwerk für Allgemeinmedizin)
EMBASE	engl. Excerpta Medica Database (Datenbank des Elsevier-Verlags für Fachartikel der Biomedizin und Pharmakologie)
EndNote	Eigenname eines Literaturverwaltungsprogramms
engl.	englisch
et al.	lat. et alii (deutsch: und andere)
etc.	lat. et cetera (deutsch: und so weiter)
FFbH-R	Funktionsfragebogen Hannover – Rückenschmerz-Version (dient der Einschätzung von Rückenschmerz-assoziierten Funktionsbeeinträchtigungen bei Alltagstätigkeiten)
Funktionskap.	Funktionskapazität
G1	Gruppe 1
G2	Gruppe 2
ggf.	gegebenenfalls
GP	engl. General Practitioner (deutsch: Allgemeinmediziner)
HA	Hausarzt
HbA <sub>1c</sub>	Hämoglobin A <sub>1c</sub> (glykiertes Hämoglobin A <sub>1</sub> )
HNO	Hals-Nasen-Ohren-Heilkunde
HWS	Halswirbelsäule
I1/I2	Interventionsgruppe 1/2
ICC	engl. Intracluster Correlation Coefficient (deutsch: Intracluster-Korrelationskoeffizient; statistisches Maß)

---

ICD-10	engl. International Statistical Classification of Diseases and Related Health Problems, version 10 (weltweit anerkanntes Diagnoseklassifikationssystem der Medizin)
ICPC-2	engl. International Classification of Primary Care, version 2 (speziell für die Allgemeinmedizin entwickeltes medizinisches Klassifikationssystem)
ID	Identifikationsnummer
IG	Interventionsgruppe
i. m.	intramuskulär
INR	engl. International Normalized Ratio (Einheit zur Messung der Blutgerinnung nach weltweitem Standard)
Interv.	Intervention
ITT	engl. Intention-to-treat (Analyseverfahren der medizinischen Statistik)
J.	Jahr(e)
K	Kontrolle
KHK	Koronare Herzkrankheit
lat.	lateinisch
LL	engl. lower level of the 95 % confidence interval (deutsch: Untergrenze des 95 % Konfidenzintervalls)
LuFu	Lungenfunktionsprüfung (hier: Spirometrie)
MEDLINE	engl. Medical Literature Analysis and Retrieval System Online (öffentlich zugängliche bibliografische Datenbank des US-amerikanischen National Center for Biotechnology Information)
MFA	Medizinische Fachangestellte
MIDAS	engl. Migraine Disability Assessment Score (Beurteilungsinstrument für durch Migräne bedingte Beeinträchtigungen)
n	Stichprobengröße/Anzahl
nb	nicht berichtet
o. ä.	oder ähnlich(e)
o. g.	oben genannt(e)
OSCE	engl. Objective Structured Clinical Examination (Prüfungsformat im Medizinstudium zur Bewertung klinisch-praktischer Fertigkeiten)
p-Wert	Signifikanzwert (Kennzahl zur Auswertung statistischer Tests)
Pat.	Patient
PDF	engl. Portable Document Format (Dateiformat für Textdateien)

PICO	engl. Patient, Intervention, Control, Outcome (Technik der evidenzbasierten Medizin zur Formulierung einer recherchierbaren Fragestellung)
PJ	Praktisches Jahr
PONV	engl. Postoperative Nausea and Vomiting (deutsch: Postoperative Übelkeit und Erbrechen)
Prob.	Proband
Prof.	Professor
PSCI	engl. Patient Safety Culture Indicator (Maß zur Beurteilung von Strukturen und Abläufen der Patientensicherheit)
PubMed	von der nationalen medizinischen Bibliothek der USA betriebene Datenbank und Suchmaschine für medizinische Fachartikel
QoL	engl. Quality of Life (deutsch: Lebensqualität)
QZ	Qualitätszirkel
Random.	Ebene der Randomisierung („klassische“ RCT vs. cRCT); <i>diese Abkürzung wurde aus Platzgründen in einigen Tabellen verwendet</i>
RCT	engl. Randomised Controlled Trial (deutsch: randomisierte kontrollierte Studie)
Schmerzintens.	Schmerzintensität
Scopus	vom Elsevier-Verlag betriebene Zitations- und Abstractdatenbank für wissenschaftliche Journalbeiträge
SDM	engl. Shared Decision Making (deutsch: partizipative Entscheidungsfindung; Form der Arzt-Patienten-Beziehung)
SF-36	engl. Short Form (36) Health Survey (Gesundheitsfragebogen zur Erhebung der gesundheitsbezogenen Lebensqualität)
SMD	engl. Standardized Mean Difference (deutsch: standardisierte Mittelwertsdifferenz)
subj.	subjektiv
Studienprot.	Studienprotokoll
tel.	telefonisch
tel.-Anb.	telefonische Anbindung
TU	Technische Universität
TUM	Technische Universität München
TV-Ä	Tarifvertrag für Ärztinnen und Ärzte
u.	und
u. g.	unten genannt(e)

UK	engl. United Kingdom (deutsch: Vereinigtes Königreich)
UL	engl. upper level oft he 95 % confidence interval (deutsch: Obergrenze des 95 % Konfidenzintervalls)
Univ.	Universität
USA	engl. United States of America (deutsch: Vereinigte Staaten von Amerika)
usw.	und so weiter
vgl.	vergleiche
vs.	lat. versus (deutsch: im Vergleich zu)
W3-Prof.	Besoldungsgruppe der deutschen Besoldungsordnung für Hochschullehrer, welcher Lehrstuhlinhaber üblicherweise angehören
Wo.	Woche(n)
z. B.	zum Beispiel



# Literaturverzeichnis

## Verzeichnis aller Text- und Studienreferenzen

Die 95 (initial) eingeschlossenen Publikationen (3.1.4 und Tabelle 32) sind fett gedruckt.

**Altiner, A., Berner, R., Diener, A., Feldmeier, G., Köchling, A., Löffler, C., Schröder, H., Siegel, A., Wollny, A., Kern, W.V.**  
***Converting habits of antibiotic prescribing for respiratory tract infections in German primary care – The cluster-randomized controlled CHANGE-2 trial.***  
**BMC Fam. Pract. 13 (2012a) 124**

**Altiner, A., Brockmann, S., Sielk, M., Wilm, S., Wegscheider, K., Abholz, H.H.**  
***Reducing antibiotic prescriptions for acute cough by motivating GPs to change their attitudes to communication and empowering patients: A cluster-randomized intervention study.***  
**J. Antimicrob. Chemother. 60 (2007) 638–644**

**Altiner, A., Schäfer, I., Mellert, C., Löffler, C., Mortsiefer, A., Ernst, A., Stolzenbach, C.O., Wiese, B., Scherer, M., van den Bussche, H., Kaduszkiewicz, H.**  
***Activating GENERAL practitioners dialogue with patients on their Agenda (MultiCare AGENDA) study protocol for a cluster randomized controlled trial.***  
**BMC Fam. Pract. 13 (2012b) 118**

BÄK (2015)  
*Ärztestatistik.*  
([http://www.bundesaerztekammer.de/fileadmin/user\\_upload/downloads/pdf-Ordner/Statistik2015/Stat15AbbTab.pdf](http://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/pdf-Ordner/Statistik2015/Stat15AbbTab.pdf))  
Letzter Zugriff: 30.01.2017

**Barzel, A., Ketels, G., Tetzlaff, B., Krüger, H., Haevernick, K., Daubmann, A., Wegscheider, K., Scherer, M.**  
***Enhancing activities of daily living of chronic stroke patients in primary health care by modified constraint-induced movement therapy (HOME CIMT): study protocol for a cluster randomized controlled trial.***  
**Trials. 14 (2013) 334**

Baum, E., Niebling, W.  
*40 Years DEGAM: General Practice at German Universities – Present State and Perspective.*  
Z. Allgemeinmed. 82 (2006) 415–419

**Becker, A., Held, H., Redaelli, M., Chenot, J.F., Leonhardt, C., Keller, S., Baum, E., Pflingsten, M., Hildebrandt, J., Basler, H.D., Kochen, M.M., Donner-Banzhoff, N., Strauch, K.**  
***Implementation of a guideline for low back pain management in primary care: A cost-effectiveness analysis.***  
**Spine (Phila. Pa. 1976). 37 (2012) 701–710**

Becker, A., Leonhardt, C., Kochen, M.M., Keller, S., Wegscheider, K., Baum, E., Donner-Banzhoff, N., Pfungsten, M., Hildebrandt, J., Basler, H.D., Chenot, J.F. ***Effects of two guideline implementation strategies on patient outcomes in primary care: A cluster randomized controlled trial.*** *Spine (Phila. Pa. 1976)*. 33 (2008) 473–480

Bergold, M., Strametz, R., Weinbrenner, S., Khan, K.S., Zamora, J., Moll, P., Weberschock, T. ***Evidence-based Medicine online for young doctors – a randomised controlled trial.*** *Z. Evid. Fortbild. Qual. Gesundh.wes.* 107 (2013) 36–43

Blank, W.A., Blankenfeld, H., Vogelmann, R., Linde, K., Schneider, A. ***Can near-peer medical students effectively teach a new curriculum in physical examination?*** *BMC Med. Educ.* 13 (2013) 165

Blank, W.A., Freiberger, E., Siegrist, M., Landendörfer, P., Linde, K., Schuster, T., Pfeifer, K., Schneider, A., Halle, M. ***An interdisciplinary intervention to prevent falls in community-dwelling elderly persons: Protocol of a cluster-randomized trial [PreFalls].*** *BMC Geriatr.* 11 (2011) 7

Bleidorn, J., Gágyor, I., Kochen, M.M., Wegscheider, K., Hummers-Pradier, E. ***Symptomatic treatment (ibuprofen) or antibiotics (ciprofloxacin) for uncomplicated urinary tract infection? – Results of a randomized controlled pilot trial.*** *BMC Med.* 8 (2010) 30

BMBF (2016)  
*Universitäten.*  
(<http://www.gesundheitsforschung-bmbf.de/de/universitaeten.php>)  
Letzter Zugriff: 30.01.2017

Borgers, D.  
*Die universitäre deutsche Allgemeinmedizin im Spiegel ihrer Zeitschriftenliteratur (1998–2007).*  
*Z. Allgemeinmed.* 84 (2008) 44–50

Bozorgmehr, K., Szecsenyi, J., Ose, D., Besier, W., Mayer, M., Krisam, J., Jacke, C.O., Salize, H.J., Brandner, R., Schmitt, S., Kiel, M., Kamradt, M., Freund, T. ***Practice network-based care management for patients with type 2 diabetes and multiple comorbidities (GEDIMaplus): study protocol for a randomized controlled trial.*** *Trials.* 15 (2014) 243

Brierley, G., Brabyn, S., Torgerson, D., Watson, J.  
*Bias in recruitment to cluster randomized trials: a review of recent publications.*  
*J. Eval. Clin. Pract.* 18 (2012) 878–886

Bücker, B., Butzlaff, M., Isfort, J., Koneczny, N., Vollmar, H.C., Lange, S., Lutz, G., Rieger, M.A.  
***Patienteninformationen und Kreuzschmerzen (PIK-Studie) – Einfluss von Patienteninformationen auf Wissen und Funktionskapazität von Patienten mit akuten, unkomplizierten Rückenschmerzen***  
*Gesundheitswesen.* 72 (2010) e78–e88

**Butzlaff, M., Vollmar, H.C., Floer, B., Koneczny, N., Isfort, J., Lange, S.**  
***Learning with computerized guidelines in general practice?: A randomized controlled trial.***  
**Fam. Pract. 21 (2004) 183–188**

Campbell, M.K., Piaggio, G., Elbourne, D.R., Altman, D.G.  
*Consort 2010 statement: extension to cluster randomised trials.*  
BMJ. 345 (2012) e5661

Chenot, J.F.  
*Cluster-randomisierte Studien: eine wichtige Methode in der allgemeinmedizinischen Forschung.*  
Z. Evid. Fortbild. Qual. Gesundh.wes. 103 (2009) 475–480

Chenot, J.F., Scherer, M., Becker, A., Donner-Banzhoff, N., Baum, E., Leonhardt, C., Keller, S., Pflingsten, M., Hildebrandt, J., Basler, H.D., Kochen, M.M.  
*Acceptance and perceived barriers of implementing a guideline for managing low back in general practice.*  
Implement. Sci. 3 (2008) 7

DAK (2009)  
*Gesundheitsreport 2009. Analyse der Arbeitsunfähigkeitsdaten. Schwerpunktthema Doping am Arbeitsplatz.*  
([https://www.dak.de/dak/download/Gesundheitsreport\\_2009\\_Doping\\_am\\_Arbeitsplatz-1117016.pdf](https://www.dak.de/dak/download/Gesundheitsreport_2009_Doping_am_Arbeitsplatz-1117016.pdf))  
Letzter Zugriff: 30.01.2017

DEGAM (2013)  
*Allgemeinmedizinische Universitätsabteilungen.*  
(<http://www.degam.de/allgemeinmedizinische-universitaetsabteilungen.html>)  
Letzter Zugriff: 30.01.2017

DEGAM-Zukunftspositionen (2012)  
*Position 20.*  
In: *Allgemeinmedizin – spezialisiert auf den ganzen Menschen, Positionen zur Zukunft der Allgemeinmedizin und der hausärztlichen Praxis.*  
Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin e. V. (Hrsg.)  
Frankfurt, 2012

Díaz-Ordaz, K., Kenward, M.G., Cohen, A., Coleman, C.L., Eldridge, S.  
*Are missing data adequately handled in cluster randomised trials? A systematic review and guidelines.*  
Clin. Trials. 11 (2014) 590–600

**Doering, T.J., Konitzer, M., Hausner, T., Steuernagel, B., Schneider, B., Fischer, G.C.**  
***Zerebrale Hämodynamik bei Kohlendioxidapplikationen.***  
**Forsch. Komplementärmed. Klass. Naturheilkd. 9 (2002) 148–152**

**Doering, T.J., Thiel, J., Steuernagel, B., Johannes, B., Konitzer, M., Niederstadt, C., Schneider, B., Fischer, G.C.**  
***Veränderung kognitiver Hirnleistungen im Alter durch Kneipp-Anwendungen.***  
**Forsch. Komplementärmed. Klass. Naturheilkd. 8 (2001) 80–84**

**Donath, C., Grassel, E., Grossfeld-Schmitz, M., Menn, P., Lauterberg, J., Wunder, S., Marx, P., Ruckdaschel, S., Mehlig, H., Holle, R.**

***Effects of general practitioner training and family support services on the care of home-dwelling dementia patients – results of a controlled cluster-randomized study.***

**BMC Health Serv. Res. 10 (2010) 314**

Donner-Banzhoff, N.

*Wagenburg und neue Kraft – Die Wissenschaft von der Allgemeinmedizin im Licht der Universitäten.*

In: *Von der allgemeinen Medizin zur Allgemeinmedizin.*

Mader, F.H. (Hrsg.)

Kirchheim-Verlag, Mainz, 2016, 50–51

**Drewelow, E., Wollny, A., Pentzek, M., Immecke, J., Lambrecht, S., Wilm, S., Schluckebier, I., Löscher, S., Wegscheider, K., Altiner, A.**

***Improvement of primary health care of patients with poorly regulated diabetes mellitus type 2 using shared decision-making – The DEBATE trial.***

**BMC Fam. Pract. 13 (2012) 88**

**Du Moulin, M., Taube, K., Wegscheider, K., Behnke, M., Van Den Bussche, H.**

***Home-based exercise training as maintenance after outpatient pulmonary rehabilitation.***

**Respiration. 77 (2009) 139–145**

Eldridge, S., Kerry, S., Torgerson, D.J.

*Bias in identifying and recruiting participants in cluster randomised trials: what can be done?*

BMJ. 339 (2009) b4006

Eldridge, S.M., Ashby, D., Feder, G.S., Rudnicka, A.R., Ukoumunne, O.C.

*Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care.*

Clinical trials. 1 (2004) 80–90

**Erler, A., Beyer, M., Petersen, J.J., Saal, K., Rath, T., Rochon, J., Haefeli, W.E., Gerlach, F.M.**

***How to improve drug dosing for patients with renal impairment in primary care – A cluster-randomized controlled trial.***

**BMC Fam. Pract. 13 (2012) 91**

Fischer, G.C.

*Erst Ablehnung, dann Anerkennung – Meine Erfahrungen als Hochschullehrerin für Allgemeinmedizin.*

In: *Von der allgemeinen Medizin zur Allgemeinmedizin.*

Mader, F.H. (Hrsg.)

Kirchheim-Verlag, Mainz, 2016, 54–55

**Freiberger, E., Blank, W.A., Salb, J., Geilhof, B., Hentschke, C., Landendörfer, P., Halle, M., Siegrist, M.**

***Effects of a complex intervention on fall risk in the general practitioner setting: A cluster randomized controlled trial.***

**Clin. Interv. Aging. 8 (2013) 1079–1088**

**Frese, T., Deutsch, T., Keyser, M., Sandholzer, H.**

***In-home preventive comprehensive geriatric assessment (CGA) reduces mortality – a randomized controlled trial.***

**Arch. Gerontol. Geriatr. 55 (2012) 639–644**

**Freund, T., Baldauf, A., Muth, C., Gensichen, J., Szecsenyi, J., Peters-Klimm, F.**

***Praxisbasiertes Hausbesuchs- und Telefonmonitoring von Patienten mit Herzinsuffizienz: Rationale, Design und praktische Anwendung der Monitoringslisten in der HICMan-Studie.***

**Z. Evid. Fortbild. Qual. Gesundh.wes. 105 (2011a) 434–445**

**Freund, T., Peters-Klimm, F., Rochon, J., Mahler, C., Gensichen, J., Erler, A., Beyer, M., Baldauf, A., Gerlach, F.M., Szecsenyi, J.**

***Primary care practice-based care management for chronically ill patients (PraCMan): Study protocol for a cluster randomized controlled trial [ISRCTN56104508].***

**Trials. 12 (2011b) 163**

**Gágyor, I., Bleidorn, J., Wegscheider, K., Hummers-Pradier, E., Kochen, M.M.**

***Practices, patients and (im)perfect data – feasibility of a randomised controlled clinical drug trial in German general practices.***

**Trials. 12 (2011) 91**

**Gágyor, I., Hummers-Pradier, E., Kochen, M.M., Schmiemann, G., Wegscheider, K., Bleidorn, J.**

***Immediate versus conditional treatment of uncomplicated urinary tract infection – a randomized-controlled comparative effectiveness study in general practices.***

**BMC Infect. Dis. 12 (2012) 146**

**Gágyor, I., Scherer, M., Bleidorn, J., Schneider, A., Hummers-Pradier, E., Beyer, M., Gerlach, F.M., Joos, S. (2014)**

***Unser Labor ist die Praxis.***

**([http://www.degam.de/files/Inhalte/Degam-](http://www.degam.de/files/Inhalte/Degam-Inhalte/Sektionen_und_Arbeitsgruppen/Sektion_Forschung/DEGAM_Broschuere_Unser_Labor_ist_die_Praxis_final.pdf)**

**[Inhalte/Sektionen\\_und\\_Arbeitsgruppen/Sektion\\_Forschung/DEGAM Broschuere\\_Unser Labor ist die Praxis\\_final.pdf](http://www.degam.de/files/Inhalte/Degam-Inhalte/Sektionen_und_Arbeitsgruppen/Sektion_Forschung/DEGAM_Broschuere_Unser_Labor_ist_die_Praxis_final.pdf))**

**Letzter Zugriff: 30.01.2017**

**Gastpar, M., Klimm, H.D.**

***Treatment of anxiety, tension and restlessness states with Kava special extract WS® 1490 in general practice: A randomized placebo-controlled double-blind multicenter trial.***

**Phytomedicine. 10 (2003) 631–639**

**Gavrylyuk, G., Ehrt, O., Meissner, K.**

***Keine Effekte von placeboinduzierten Erwartungen auf die Pupillengröße und die Augenakkommodation.***

**Z. Med. Psychol. 19 (2010) 154–160**

Gensichen, J., Hiller, T.S., Breitbart, J., Teismann, T., Brettschneider, C., Schumacher, U., Piwitorak, A., König, H.H., Hoyer, H., Schneider, N., Schelle, M., Blank, W., Thiel, P., Wensing, M., Margraf, J.  
*Evaluation of a practice team-supported exposure training for patients with panic disorder with or without agoraphobia in primary care – study protocol of a cluster randomised controlled superiority trial.*

*Trials.* 15 (2014) 112

Gensichen, J., Petersen, J.J., Karroum, T., Rauck, S., Ludman, E., König, J., Gerlach, F.M.

*Positive impact of a family practice-based depression case management on patient's self-management.*

*Gen. Hosp. Psychiatry.* 33 (2011) 23–28

Gensichen, J., Torge, M., Peitz, M., Wendt-Hermainski, H., Beyer, M., Rosemann, T., Krauth, C., Raspe, H., Aldenhoff, J.B., Gerlach, F.M.

*Case management for the treatment of patients with major depression in general practices – Rationale, design and conduct of a cluster randomized controlled trial – PRoMPT (Primary care Monitoring for depressive Patient's Trial) [ISRCTN66386086] – Study protocol.*

*BMC Public Health.* 5 (2005) 101

Gensichen, J., Von Korff, M., Peitz, M., Muth, C., Beyer, M., Gütthlin, C., Torge, M., Petersen, J.J., Rosemann, T., König, J., Gerlach, F.M.

*Case management for depression by health care assistants in small primary care practices: A cluster randomized trial.*

*Ann. Intern. Med.* 151 (2009) 369–378

Gerlach, F.M.

*Vorwort.*

In: *Von der allgemeinen Medizin zur Allgemeinmedizin.*

Mader, F.H. (Hrsg.)

Kirchheim-Verlag, Mainz, 2016, 4

Glanville, J., Kendrick, T., McNally, R., Campbell, J., Hobbs, F.R.

*Research output on primary care in Australia, Canada, Germany, the Netherlands, the United Kingdom, and the United States: bibliometric analysis.*

*BMJ.* 342 (2011) d1028

Grätzel von Grätz, P.

*Plädoyer für ein hausärztliches Forschungsnetz.*

*Der Hausarzt.* 7 (2015) 34–35

Green, L.A., Fryer, G.E., Jr., Yawn, B.P., Lanier, D., Dovey, S.M.

*The ecology of medical care revisited.*

*N. Engl. J. Med.* 344 (2001) 2021–2025

Grossmann, N.

*Publikation von Forschungsarbeiten durch die universitäre deutsche Allgemeinmedizin – eine systematische Übersicht von 2000 bis 2010.*

Technische Universität München. (2014)

Grossmann, N., Schneider, A., Linde, K.

*Publikation von Forschungsarbeiten durch die deutsche akademische Allgemeinmedizin von 2000 bis 2010.*

Z. Allgemeinmed. 88 (2012) 345–354

**Gummersbach, E., In der Schmitt, J., Abholz, H.H., Wegscheider, K., Pentzek, M.**  
***Effects of different information brochures on women's decision-making regarding mammography screening: Study protocol for a randomized controlled questionnaire study.***

**Trials. 14 (2013) 319**

Güthlin, C., Beyer, M., Eler, A., Gensichen, J., Hoffmann, B., Mergenthal, K., Müller, V., Muth, C., Petersen, J.J., Gerlach, F.M.

*Rekrutierung von Hausarztpraxen für Forschungsprojekte: Erfahrungen aus fünf allgemeinmedizinischen Studien.*

Z. Allgemeinmed. 88 (2012) 173–181

Habilitationen

*Medizinische Fakultäten: Habilitationen im 1. Quartal 2015.*

Dtsch. Ärztebl. Int. 112 (2015a) A-1296 / B-1086 / C-1058

Habilitationen

*Medizinische Fakultäten: Habilitationen im 2. Quartal 2015.*

Dtsch. Ärztebl. Int. 112 (2015b) A-1584 / B-1318 / C-1288

Habilitationen

*Medizinische Fakultäten: Habilitationen im 3. Quartal 2015.*

Dtsch. Ärztebl. Int. 112 (2015c) A-1952 / B-1612 / C-1564

Habilitationen

*Medizinische Fakultäten: Habilitationen im 4. Quartal 2015.*

Dtsch. Ärztebl. Int. 113 (2015d) A-386 / B-328 / C-328

Häussler, S.

*Das Heidelberger Gespräch zwischen Klinikern und Praktikern.*

In: *Von der allgemeinen Medizin zur Allgemeinmedizin.*

Mader, F.H. (Hrsg.)

Kirchheim-Verlag, Mainz, 2016, 46–48

Hedges, L.V., Olkin, I.

*Vote-counting methods in research synthesis.*

Psychol. Bull. 88 (1980) 359–369

Heinmüller, S., Schneider, A., Linde, K.

*Quantity, topics, methods and findings of randomised controlled trials published by German university departments of general practice – systematic review.*

Trials. 17 (2016a) 211

Heinmüller, S., Schneider, A., Linde, K.

*Randomisierte Studien der deutschen universitären Allgemeinmedizin – Ergebnisse einer systematischen Übersichtsarbeit.*

Z. Allgemeinmed. 92 (2016b) 259–264

**Hensler, S., Guendling, P.W., Schmidt, M., Jork, K.**  
***Autologous blood therapy for common cold – A randomized, double-blind, placebo-controlled trial.***  
**Complement. Ther. Med. 17 (2009) 257–261**

Higgins, J., Green, S. (2011)  
*Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 [updated March 2011].*  
(<http://www.handbook.cochrane.org/>)  
Letzter Zugriff: 30.01.2017

**Hirsch, O., Keller, H., Albohn-Kühne, C., Krones, T., Donner-Banzhoff, N.**  
***Satisfaction of patients and primary care physicians with shared decision making.***  
**Eval. Health Prof. 33 (2010) 321–342**

**Hoffmann, B., Müller, V., Rochon, J., Gondan, M., Müller, B., Albay, Z., Wepler, K., Leifermann, M., Miessner, C., Güthlin, C., Parker, D., Hofinger, G., Gerlach, F.M.**  
***Effects of a team-based assessment and intervention on patient safety culture in general practice: an open randomised controlled trial.***  
**BMJ Qual. Saf. 23 (2014) 35–46**

**Holle, R., Gräbel, E., Ruckdäschel, S., Wunder, S., Mehlig, H., Marx, P., Pirk, O., Butzlaff, M., Kunz, S., Lauterberg, J.**  
***Dementia care initiative in primary practice study protocol of a cluster randomized trial on dementia management in a general practice setting.***  
**BMC Health Serv. Res. 9 (2009) 91**

**Hua, T.D., Vormfelde, S.V., Abed, M.A., Schneider-Rudt, H., Sobotta, P., Friede, T., Chenot, J.F.**  
***Practice nursed-based, individual and video-assisted patient education in oral anticoagulation – Protocol of a cluster-randomized controlled trial.***  
**BMC Fam. Pract. 12 (2011) 17**

Hummers-Pradier, E.  
*40 Years DEGAM: Research in general practice in Germany – Development and current situation.*  
Z. Allgemeinmed. 82 (2006) 409–414

Hummers-Pradier, E., Beyer, M., Chevallier, P., Eilat-Tsanani, S., Lionis, C., Peremans, L., Petek, D., Rurik, I., Soler, J.K., Stoffers, H.E., Topsever, P., Ungan, M., Van Royen, P. (2009)  
*The Research Agenda for General Practice/Family Medicine and Primary Health Care in Europe.*  
([http://www.egprn.org/files/userfiles/file/research\\_agenda\\_for\\_general\\_practice\\_family\\_medicine.pdf](http://www.egprn.org/files/userfiles/file/research_agenda_for_general_practice_family_medicine.pdf))  
Letzter Zugriff: 30.01.2017

Hummers-Pradier, E., Bleidorn, J., Schmiemann, G., Joos, S., Becker, A., Altiner, A., Chenot, J.F., Scherer, M.  
*General practice-based clinical trials in Germany – a problem analysis.*  
Trials. 13 (2012) 205

Hummers-Pradier, E., Chenot, J.F., Scherer, M.  
*Are primary care practices research infrastructure?*  
Z. Allgemeinmed. 90 (2014) 317–322

ICD-10 (2016)  
*Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme, 10. Revision (German Modification).*  
(<http://www.icd-code.de/icd/code/ICD-10-GM-2016.html>)  
Letzter Zugriff: 30.01.2017

ICPC-2 (2011)  
*International Classification of Primary Care – 2nd Edition.*  
(<http://www.kith.no/upload/2705/icpc-2-german.pdf>)  
Letzter Zugriff: 30.01.2017

IQWiG (2016)  
*Glossar zu Begriffen aus dem Gesundheitswesen.*  
([https://www.iqwig.de/download/IQWiG\\_Glossar.pdf](https://www.iqwig.de/download/IQWiG_Glossar.pdf))  
Letzter Zugriff: 30.01.2017

**Jäger, C., Freund, T., Steinhäuser, J., Joos, S., Wensing, M., Szecsenyi, J.**  
***A tailored implementation intervention to implement recommendations addressing polypharmacy in multimorbid patients: Study protocol of a cluster randomized controlled trial.***  
Trials. 14 (2013) 420

**Jobst, D., Altiner, A., Wegscheider, K., Abholz, H.H.**  
***Helfen intramuskuläre Eigenblutgaben bei chronisch rezidivierenden Infekten der Atemwege? – Fußangeln auf dem Weg einer randomisierten Studie.***  
Z. Allgemeinmed. 81 (2005) 258–263

**Joos, S., Brinkhaus, B., Maluche, C., Maupai, N., Kohnen, R., Hahn, E., Schuppan, D.**  
***Akupunktur und Moxibustion als ergänzende Therapie bei M. Crohn.***  
Dtsch. Z. Akup. 48 (2005a) 6–17

**Joos, S., Brinkhaus, B., Maluche, C., Maupai, N., Kohnen, R., Kraehmer, N., Hahn, E.G., Schuppan, D.**  
***Acupuncture and moxibustion in the treatment of active Crohn's disease: a randomized controlled study.***  
Digestion. 69 (2004) 131–139

Joos, S., Musselmann, B., Szecsenyi, J.  
*Integration of complementary and alternative medicine into family practices in Germany: results of a national survey.*  
Evid. Based Complement. Alternat. Med. (2011) Article-ID 495813

**Joos, S., Rosemann, T., Heiderhoff, M., Wensing, M., Ludt, S., Gensichen, J., Kaufmann-Kolle, P., Szecsenyi, J.**  
***ELSID-Diabetes study-evaluation of a large scale implementation of disease management programmes for patients with type 2 diabetes. Rationale, design and conduct – A study protocol [ISRCTN08471887].***  
BMC Public Health. 5 (2005b) 99

**Joos, S., Wildau, N., Kohnen, R., Szecsenyi, J., Schuppan, D., Willich, S.N., Hahn, E.G., Brinkhaus, B.**  
***Acupuncture and moxibustion in the treatment of ulcerative colitis: a randomized controlled study.***  
**Scand. J. Gastroenterol. 41 (2006) 1056–1063**

Josupeit, T.  
*Bewertung der akademischen Allgemeinmedizin – Eine schriftliche Befragung von Hochschulprofessoren an den deutschen medizinischen Fakultäten.*  
Universität Göttingen. (2002)

**Junius-Walker, U., Wrede, J., Voigt, I., Hofmann, W., Wiese, B., Hummers-Pradier, E., Dierks, M.L.**  
***Impact of a priority-setting consultation on doctor-patient agreement after a geriatric assessment: Cluster randomised controlled trial in German general practices.***  
**Qual. Prim. Care. 20 (2012) 321–334**

Kabisch, M., Ruckes, C., Seibert-Grafe, M., Blettner, M.  
*Randomized controlled trials: part 17 of a series on evaluation of scientific publications.*  
Dtsch. Ärztebl. Int. 108 (2011) 663–668

**Katon, W.J., Lin, E.H., Von Korff, M., Ciechanowski, P., Ludman, E.J., Young, B., Peterson, D., Rutter, C.M., McGregor, M., McCulloch, D.**  
***Collaborative care for patients with depression and chronic illnesses.***  
**N. Engl. J. Med. 363 (2010) 2611–2620**

**Kaufmann-Kolle, P., Szecsenyi, J., Broge, B., Haefeli, W.E., Schneider, A.**  
***Führt die Implementierung von offenem Benchmarking in datengestützten Qualitätszirkeln zur Verbesserung der hausärztlichen Versorgung bei Arzneimittelinteraktion und Asthma bronchiale?***  
**Z. Evid. Fortbild. Qual. Gesundh.wes. 105 (2011) 389–395**

KBV (2016)  
*Ärztemangel.*  
([http://www.kbv.de/html/themen\\_1076.php](http://www.kbv.de/html/themen_1076.php))  
Letzter Zugriff: 30.01.2017

**Keller, H., Hirsch, O., Kaufmann-Kolle, P., Krones, T., Becker, A., Sönnichsen, A.C., Baum, E., Donner-Banzhoff, N.**  
***Evaluating an implementation strategy in cardiovascular prevention to improve prescribing of statins in Germany: An intention to treat analysis.***  
**BMC Public Health. 13 (2013) 623**

**Keller, H., Krones, T., Becker, A., Hirsch, O., Sönnichsen, A.C., Popert, U., Kaufmann-Kolle, P., Rochon, J., Wegscheider, K., Baum, E., Donner-Banzhoff, N.**  
***Arriba: Effects of an educational intervention on prescribing behaviour in prevention of CVD in general practice.***  
**Eur. J. Prev. Cardiol. 19 (2012) 322–329**

**Klein, R., Bareis, A., Schneider, A., Linde, K.**  
***Strain-counterstrain to treat restrictions of the mobility of the cervical spine in patients with neck pain – A sham-controlled randomized trial.***  
**Complement. Ther. Med. 21 (2013) 1–7**

Kochen, M.M.

*Editorial: New impulses for general practice in Germany.*

Eur. J. Gen. Pract. 5 (1999) 10

Kortekaas, M., Meijer, A., van de Pol, A., de Wit, N.

*RCT's in de huisartsgeneeskunde, 1990-2010.*

Huisarts en wetenschap. 55 (2012) 486–491

**Krones, T., Keller, H., Becker, A., Sönnichsen, A., Baum, E., Donner-Banzhoff, N.**  
***The theory of planned behaviour in a randomized trial of a decision aid on cardiovascular risk prevention.***

Patient Educ. Couns. 78 (2010) 169–176

**Krones, T., Keller, H., Sönnichsen, A., Sadowski, E.M., Baum, E., Wegscheider, K., Rochon, J., Donner-Banzhoff, N.**

***Absolute cardiovascular disease risk and shared decision making in primary care: A randomized controlled trial.***

Ann. Fam. Med. 6 (2008) 218–227

**Lauterberg, J., Großfeld-Schmitz, M., Ruckdäschel, S., Neubauer, S., Mehlig, H., Gaudig, M., Hruschka, D., Vollmar, H.C., Holle, R., Gräbel, E.**

***Projekt IDA – Konzept und Umsetzung einer cluster-randomisierten Studie zur Demenzversorgung im hausärztlichen Bereich.***

Z. Ärztl. Fortbild. Qual. Gesundh.wes. 101 (2007) 21–26

**Leonhardt, C., Keller, S., Chenot, J.F., Luckmann, J., Basler, H.D., Wegscheider, K., Baum, E., Donner-Banzhoff, N., Pfingsten, M., Hildebrandt, J., Kochen, M.M., Becker, A.**

***TTM-based motivational counselling does not increase physical activity of low back pain patients in a primary care setting – A cluster-randomized controlled trial.***

Patient Educ. Couns. 70 (2008) 50–60

**Lin, E.H., Von Korff, M., Ciechanowski, P., Peterson, D., Ludman, E.J., Rutter, C.M., Oliver, M., Young, B.A., Gensichen, J., McGregor, M., McCulloch, D.K., Wagner, E.H., Katon, W.J.**

***Treatment adjustment and medication adherence for complex patients with diabetes, heart disease, and depression: a randomized controlled trial.***

Ann. Fam. Med. 10 (2012) 6–14

**Linde, K., Alscher, A., Friedrichs, C., Wagenpfeil, S., Karsch-Völk, M., Schneider, A.**

***Belief in and use of complementary therapies among family physicians, internists and orthopaedists in Germany – cross-sectional survey.***

Fam. Pract. 32 (2015) 62–68

**Löffler, C., Böhmer, F., Hornung, A., Lang, H., Burmeister, U., Podbielski, A., Wollny, A., Kundt, G., Altiner, A.**

***Dental care resistance prevention and antibiotic prescribing modification – the cluster-randomised controlled DREAM trial.***

Implement. Sci. 9 (2014a) 27

**Löffler, C., Drewelow, E., Paschka, S.D., Frankenstein, M., Eger, J., Jatsch, L., Reisinger, E.C., Hallauer, J.F., Drewelow, B., Heidorn, K., Schröder, H., Wollny, A., Kundt, G., Schmidt, C., Altiner, A.**  
***Optimizing polypharmacy among elderly hospital patients with chronic diseases – study protocol of the cluster randomized controlled POLITE-RCT trial.***  
**Implement. Sci. 9 (2014b) 151**

**Ludman, E.J., Peterson, D., Katon, W.J., Lin, E.H., Von Korff, M., Ciechanowski, P., Young, B., Gensichen, J.**  
***Improving confidence for self care in patients with depression and chronic illnesses.***  
**Behav. Med. 39 (2013) 1–6**

Mader, F.H.  
*Allgemeinmedizin und Praxis.*  
Springer-Verlag, Berlin/Heidelberg, 2014

Mader, F.H.  
*Einleitung.*  
In: *Von der allgemeinen Medizin zur Allgemeinmedizin.*  
Mader, F.H. (Hrsg.)  
Kirchheim-Verlag, Mainz, 2016, 5

Maier, W., Linden, M., Sartorius, N.  
*Psychische Erkrankungen in der Allgemeinpraxis: Ergebnisse und Schlussfolgerungen einer WHO-Studie.*  
Dtsch. Ärztebl. Int. 93 (1996) 1202–1206

**Mehring, M., Haag, M., Linde, K., Wagenpfeil, S., Frensch, F., Blome, J., Schneider, A.**  
***Effects of a general practice guided web-based weight reduction program – Results of a cluster-randomized controlled trial.***  
**BMC Fam. Pract. 14 (2013) 76**

**Mehring, M., Haag, M., Linde, K., Wagenpfeil, S., Schneider, A.**  
***Effects of a guided web-based smoking cessation program with telephone counseling: a cluster randomized controlled trial.***  
**J. Med. Internet Res. 16 (2014) e218**

**Meissner, K., Ziep, D.**  
***Organ-specificity of placebo effects on blood pressure.***  
**Auton. Neurosci. 164 (2011) 62–66**

Melchart, D., Linde, K., Fischer, P., White, A., Allais, G., Vickers, A., Berman, B.  
*Acupuncture for recurrent headaches: a systematic review of randomized controlled trials.*  
Cephalalgia. 19 (1999) 779–786

**Menn, P., Holle, R., Kunz, S., Donath, C., Lauterberg, J., Leidl, R., Marx, P., Mehlig, H., Ruckdäschel, S., Vollmar, H.C., Wunder, S., Gräbel, E.**  
***Dementia care in the general practice setting: A cluster randomized trial on the effectiveness and cost impact of three management strategies.***  
**Value Health. 15 (2012) 851–859**

Mortsiefer, A., Meysen, T., Schumacher, M., Lintges, C., Stamer, M., Schmacke, N., Wegscheider, K., Abholz, H.H., In Der Schmitten, J.  
***CRISTOPH – A cluster-randomised intervention study to optimise the treatment of patients with hypertension in General Practice.***  
BMC Fam. Pract. 9 (2008) 33

Müller-Bühl, U., Franke, B., Hermann, K., Engeser, P.  
***Lowering missing item values in quality-of-life questionnaires: an interventional study.***  
Int. J. Public Health. 56 (2011) 63–69

Peters-Klimm, F., Campbell, S., Hermann, K., Kunz, C.U., Müller-Tasch, T., Szecsenyi, J.  
***Case management for patients with chronic systolic heart failure in primary care: The HICMan exploratory randomised controlled trial.***  
Trials. 11 (2010a) 56

Peters-Klimm, F., Campbell, S., Müller-Tasch, T., Schellberg, D., Gelbrich, G., Herzog, W., Szecsenyi, J.  
***Primary care-based multifaceted, interdisciplinary medical educational intervention for patients with systolic heart failure: Lessons learned from a cluster randomised controlled trial.***  
Trials. 10 (2009) 68

Peters-Klimm, F., Kunz, C.U., Laux, G., Szecsenyi, J., Müller-Tasch, T.  
***Patient- and provider-related determinants of generic and specific health-related quality of life of patients with chronic systolic heart failure in primary care: A cross-sectional study.***  
Health Qual. Life Out. 8 (2010b) 98

Peters-Klimm, F., Laux, G., Campbell, S., Müller-Tasch, T., Lossnitzer, N., Schultz, J.H., Remppis, A., Jünger, J., Nikendei, C.  
***Physician and patient predictors of evidence-based prescribing in heart failure: A multilevel study.***  
PLoS One. 7 (2012) e31082

Peters-Klimm, F., Müller-Tasch, T., Remppis, A., Szecsenyi, J., Schellberg, D.  
***Improved guideline adherence to pharmacotherapy of chronic systolic heart failure in general practice – Results from a cluster-randomized controlled trial of implementation of a clinical practice guideline.***  
J. Eval. Clin. Pract. 14 (2008) 823–829

Peters-Klimm, F., Müller-Tasch, T., Schellberg, D., Gensichen, J., Muth, C., Herzog, W., Szecsenyi, J.  
***Rationale, design and conduct of a randomised controlled trial evaluating a primary care-based complex intervention to improve the quality of life of heart failure patients: HICMan (Heidelberg Integrated Case Management).***  
BMC Cardiovasc. Disord. 7 (2007) 25

Petersen, J.J., König, J., Paulitsch, M.A., Mergenthal, K., Rauck, S., Pagitz, M., Schmidt, K., Haase, L., Gerlach, F.M., Gensichen, J.  
***Long-term effects of a collaborative care intervention on process of care in family practices in Germany: a 24-month follow-up study of a cluster randomized controlled trial.***  
Gen. Hosp. Psychiatry. 36 (2014) 570–574

PICO (2016)  
*PICO-Modell.*  
(<https://de.wikipedia.org/wiki/PICO>)  
Letzter Zugriff: 30.01.2017

Richter-Kuhlmann, E., Rieser, S.  
*Allgemeinmedizin: Ein Fach im Aufwind.*  
Dtsch. Ärztebl. Int. 112 (2015) 1866–1870

Ronel, J., Mehilli, J., Ladwig, K.H., Blättler, H., Oversohl, N., Byrne, R.A., Bauer, A., Schneider, S., Linde, K., Henningsen, P., Lahmann, C., Noll-Hussong, M., Meissner, K.  
***Effects of verbal suggestion on coronary arteries: Results of a randomized controlled experimental investigation during coronary angiography.***  
Am. Heart J. 162 (2011) 507–511

Rosemann, T., Joos, S., Laux, G., Gensichen, J., Szecsenyi, J.  
***Case management of arthritis patients in primary care: A cluster-randomized controlled trial.***  
Arthritis Care Res. 57 (2007) 1390–1397

Rosemann, T., Körner, T., Wensing, M., Gensichen, J., Muth, C., Joos, S., Szecsenyi, J.  
***Rationale, design and conduct of a comprehensive evaluation of a primary care based intervention to improve the quality of life of osteoarthritis patients. The PraxArt-project: A cluster randomized controlled trial [ISRCTN87252339].***  
BMC Public Health. 5 (2005) 77

Schencking, M., Otto, A., Deutsch, T., Sandholzer, H.  
***A comparison of Kneipp hydrotherapy with conventional physiotherapy in the treatment of osteoarthritis of the hip or knee: Protocol of a prospective randomised controlled clinical trial.***  
BMC Musculoskelet. Disord. 10 (2009) 104

Schencking, M., Wilm, S., Redaelli, M.  
***A comparison of Kneipp hydrotherapy with conventional physiotherapy in the treatment of osteoarthritis: A pilot trial.***  
J. Integr. Med. 11 (2013) 17–25

Schmidt, C.O., Pflingsten, M., Fahland, R.A., Lindena, G., Marnitz, U., Pfeifer, K., Kohlmann, T., Chenot, J.F.  
***Assessing a risk tailored intervention to prevent disabling low back pain – Protocol of a cluster randomized controlled trial.***  
BMC Musculoskelet. Disord. 11 (2010) 5

Schmidt, K., Thiel, P., Mueller, F., Schmuecker, K., Worrack, S., Mehlhorn, J., Engel, C., Brenk-Franz, K., Kausche, S., Jakobi, U., Bindara-Klippel, A., Schneider, N., Freytag, A., Davydow, D., Wensing, M., Brunkhorst, F.M., Gensichen, J.  
***Sepsis survivors monitoring and coordination in outpatient health care (SMOOTH): study protocol for a randomized controlled trial.***  
Trials. 15 (2014) 283

Schmittdiel, M.C.  
*Innovative Lösungen bei der sektorenübergreifenden Kooperation, dargestellt am Beispiel der ambulanten Notfallversorgung in der Bereitschaftspraxis Münchner Ärzte im Klinikum Rechts der Isar der TU München.*  
Technische Universität München. (2002)

Schneider, A.  
*Die Entwicklung der akademischen Allgemeinmedizin – Erfolgreiche Förderung durch das Bildungs- und Forschungsministerium.*  
In: *Von der allgemeinen Medizin zur Allgemeinmedizin.*  
Mader, F.H. (Hrsg.)  
Kirchheim-Verlag, Mainz, 2016, 56–57

Schneider, A., Enck, P., Streitberger, K., Joos, S., Weiland, C., Bagheri, S., Friederich, H.C., Zipfel, S.  
***Spezifische physiologische und unspezifische psychische Effekte bei der Akupunkturtherapie des Reizdarmsyndroms: Ergebnisse einer randomisierten kontrollierten Studie.***  
Dtsch. Z. Akup. 51 (2008a) 8–16

Schneider, A., Enck, P., Streitberger, K., Weiland, C., Bagheri, S., Witte, S., Friederich, H.C., Herzog, W., Zipfel, S.  
***Acupuncture treatment in irritable bowel syndrome.***  
Gut. 55 (2006) 649–654

Schneider, A., Grossmann, N., Linde, K.  
*The development of general practice as an academic discipline in Germany – an analysis of research output between 2000 and 2010.*  
BMC Fam. Pract. 13 (2012) 58

Schneider, A., Löwe, B., Streitberger, K.  
***Perception of bodily sensation as a predictor of treatment response to acupuncture for postoperative nausea and vomiting prophylaxis.***  
J. Altern. Complement. Med. 11 (2005) 119–125

Schneider, A., Weiland, C., Enck, P., Joos, S., Streitberger, K., Maser-Gluth, C., Zipfel, S., Bagheri, S., Herzog, W., Friederich, H.C.  
***Neuroendocrinological effects of acupuncture treatment in patients with irritable bowel syndrome.***  
Complement. Ther. Med. 15 (2007) 255–263

Schneider, A., Wensing, M., Biessecker, K., Quinzler, R., Kaufmann-Kolle, P., Szecsenyi, J.  
***Impact of quality circles for improvement of asthma care: Results of a randomized controlled trial.***  
J. Eval. Clin. Pract. 14 (2008b) 185–190

Schünemann, H., Oxman, A., Vist, G., Higgins, J., Deeks, J., Glasziou, P., Guyatt, G. (2011)  
*Chapter 12: Interpreting results and drawing conclusions.*  
In: *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011).* The Cochrane Collaboration, 2011.  
Higgins J., Green S. (editors)  
(<http://www.handbook.cochrane.org/>)  
Letzter Zugriff: 30.01.2017

Schwartz, F.W.  
*Forschung und Allgemeinmedizin: Grundlagen und Forderungen.*  
Der Praktische Arzt. 9 (1983) 665–673

Shi, L.  
*Primary care, specialty care, and life chances.*  
Int. J. Health Serv. 24 (1994) 431–458

**Siebenhofer, A., Ulrich, L.R., Mergenthal, K., Roehl, I., Rauck, S., Berghold, A., Harder, S., Gerlach, F.M., Petersen, J.J.**  
***Primary care management for optimized antithrombotic treatment [PICANT]: study protocol for a cluster-randomized controlled trial.***  
Implement. Sci. 7 (2012) 79

Starfield, B.  
*Is primary care essential?*  
Lancet. 344 (1994) 1129–1133

**Streitberger, K., Diefenbacher, M., Bauer, A., Conradi, R., Bardenheuer, H., Martin, E., Schneider, A., Unnebrink, K.**  
***Acupuncture compared to placebo-acupuncture for postoperative nausea and vomiting prophylaxis: A randomised placebo-controlled patient and observer blind trial.***  
Anaesthesia. 59 (2004) 142–149

Szecsényi, J., Miksch, A., Ose, D., Heiderhoff, M., Rölz, A., Trieschmann, J., Riens, B., Kaufmann-Kolle, P., Broge, B. (2011)  
***Evaluation des DMP Diabetes mellitus Typ 2 im Rahmen der ELSID-Studie, Abschlussbericht für den AOK-Bundesverband.***  
([http://www.aok-gesundheitspartner.de/imperia/md/gpp/bund/dmp/evaluation/elsid/dmp\\_elsid\\_abschlussbericht\\_2012.pdf](http://www.aok-gesundheitspartner.de/imperia/md/gpp/bund/dmp/evaluation/elsid/dmp_elsid_abschlussbericht_2012.pdf))  
Letzter Zugriff: 30.01.2017

Tinsel, I., Buchholz, A., Vach, W., Siegel, A., Dürk, T., Buchholz, A., Niebling, W., Fischer, K.G.  
***Shared decision-making in antihypertensive therapy: A cluster randomised controlled trial.***  
BMC Fam. Pract. 14 (2013) 135

Tinsel, I., Buchholz, A., Vach, W., Siegel, A., Dürk, T., Loh, A., Buchholz, A., Niebling, W., Fischer, K.G.  
***Implementation of shared decision making by physician training to optimise hypertension treatment. Study protocol of a cluster-RCT.***  
BMC Cardiovasc. Disord. 12 (2012) 73

- Voigt, I., Wrede, J., Diederichs-Egidi, H., Dierks, M.L., Hummers-Pradier, E., Junius-Walker, U.  
*PräfCheck: Patientenzentrierte Behandlungsplanung mit älteren multimorbiden Patienten.*  
Z. Gerontol. Geriatr. 43 (2010) 303–309
- Voigt, K., Liebnitzky, J., Burmeister, U., Sihvonen-Riemenschneider, H., Beck, M., Voigt, R., Bergmann, A.  
*Efficacy of osteopathic manipulative treatment of female patients with migraine: Results of a randomized controlled trial.*  
J. Altern. Complement. Med. 17 (2011) 225–230
- Vollmar, H.C., Butzlaff, M.E., Lefering, R., Rieger, M.A.  
*Knowledge translation on dementia: A cluster randomized trial to compare a blended learning approach with a "classical" advanced training in GP quality circles.*  
BMC Health Serv. Res. 7 (2007a) 92
- Vollmar, H.C., Gräbel, E., Lauterberg, J., Neubauer, S., Großfeld-Schmitz, M., Koneczny, N., Schürer-Maly, C.C., Koch, M., Ehlert, N., Holle, R., Rieger, M.A., Butzlaff, M.  
*Multimodale Schulung von Hausärzten – Evaluation und Wissenszuwachs im Rahmen der Initiative Demenzversorgung in der Allgemeinmedizin (IDA).*  
Z. Evid. Fortbild. Qual. Gesundh.wes. 101 (2007b) 27–34
- Vollmar, H.C., Mayer, H., Ostermann, T., Butzlaff, M.E., Sandars, J.E., Wilm, S., Rieger, M.A.  
*Knowledge transfer for the management of dementia: A cluster-randomised trial of blended learning in general practice.*  
Implement. Sci. 5 (2010a) 1
- Vollmar, H.C., Mayer, H., Rieger, M., Wilm, S., Ostermann, T.  
*Blended Learning zur hausärztlichen Demenz-Fortbildung: Einschätzung durch Nutzer einer cluster-randomisierten Studie.*  
GMS Med. Inform. Biom. Epidemiol. 6 (2010b) 1–8
- Vormfelde, S.V., Abu Abed, M., Hua, T.D., Schneider, S., Friede, T., Chenot, J.F.  
*Educating orally anticoagulated patients in drug safety: a cluster-randomized study in general practice.*  
Dtsch. Ärztebl. Int. 111 (2014) 607–614
- Wrede, J., Voigt, I., Bleidorn, J., Hummers-Pradier, E., Dierks, M.L., Junius-Walker, U.  
*Complex health care decisions with older patients in general practice: Patient-centeredness and prioritization in consultations following a geriatric assessment.*  
Patient Educ. Couns. 90 (2013) 54–60
- Zimmermann, T., Puschmann, E., Ebersbach, M., Daubmann, A., Steinmann, S., Scherer, M.  
*Effectiveness of a primary care based complex intervention to promote self-management in patients presenting psychiatric symptoms: study protocol of a cluster-randomized controlled trial.*  
BMC Psychiatry. 14 (2014) 2

## Übersicht aller Studienreferenzen

Tabelle 32 schlüsselt die 95 Publikationen (3.1.4) nach ihrer Studienzugehörigkeit auf. In Text/Tabellen genannte Studieninformationen stammen aus den hier genannten Artikeln.

Tabelle 32: Aufschlüsselung der 95 Publikationen nach ihrer Zugehörigkeit zu den 60 RCTs

<p><b>Altiner 2007</b> Altiner et al. [2007]</p> <p><b>Altiner 2012a</b> Altiner et al. [2012a] (P)</p> <p><b>Altiner 2012b</b> Altiner et al. [2012b] (P)</p> <p><b>Barzel 2013</b> Barzel et al. [2013] (P)</p> <p><b>Becker 2008</b> *Becker et al. [2008] Becker et al. [2012] Leonhardt et al. [2008]</p> <p><b>Bergold 2013</b> Bergold et al. [2013]</p> <p><b>Blank 2013</b> Blank et al. [2013]</p> <p><b>Bleidorn 2010</b> *Bleidorn et al. [2010] Gágyor et al. [2011]</p> <p><b>Bozorgmehr 2014</b> Bozorgmehr et al. [2014] (P)</p> <p><b>Bücker 2010</b> Bücker et al. [2010]</p> <p><b>Butzlaff 2004</b> Butzlaff et al. [2004]</p> <p><b>Doering 2001</b> Doering et al. [2001]</p> <p><b>Doering 2002</b> Doering et al. [2002]</p> <p><b>Drewelow 2012</b> Drewelow et al. [2012] (P)</p> <p><b>Du Moulin 2009</b> Du Moulin et al. [2009]</p> <p><b>Erler 2012</b> Erler et al. [2012]</p> <p><b>Freiberger 2013</b> **Blank et al. [2011] (P) *Freiberger et al. [2013]</p> <p><b>Frese 2012</b> Frese et al. [2012]</p> <p><b>Freund 2011</b> Freund et al. [2011b] (P)</p> <p><b>Gágyor 2012</b> Gágyor et al. [2012] (P)</p> <p><b>Gastpar 2003</b> Gastpar et al. [2003]</p> <p><b>Gavrylyuk 2010</b> Gavrylyuk et al. [2010]</p> <p><b>Gensichen 2009</b> Gensichen et al. [2005] (P) *Gensichen et al. [2009] Gensichen et al. [2011] Petersen et al. [2014]</p>	<p><b>Gensichen 2014</b> Gensichen et al. [2014] (P)</p> <p><b>Gummersbach 2013</b> Gummersbach et al. [2013] (P)</p> <p><b>Hensler 2009</b> Hensler et al. [2009]</p> <p><b>Hoffmann 2014</b> Hoffmann et al. [2014]</p> <p><b>Jäger 2013</b> Jäger et al. [2013] (P)</p> <p><b>Jobst 2005</b> Jobst et al. [2005]</p> <p><b>Joos 2005</b> *Joos et al. [2004] **Joos et al. [2005a]</p> <p><b>Joos 2006</b> Joos et al. [2006]</p> <p><b>Junius-Walker 2012</b> *Junius-Walker et al. [2012] Voigt et al. [2010] (P) Wrede et al. [2013]</p> <p><b>Kaufmann-Kolle 2011</b> *Kaufmann-Kolle et al. [2011] Schneider et al. [2008b]</p> <p><b>Klein 2013</b> Klein et al. [2013]</p> <p><b>Krones 2008</b> Hirsch et al. [2010] Keller et al. [2012] Keller et al. [2013] *Krones et al. [2008] Krones et al. [2010]</p> <p><b>Löffler 2014a</b> Löffler et al. [2014a] (P)</p> <p><b>Löffler 2014b</b> Löffler et al. [2014b] (P)</p> <p><b>Ludman 2013</b> *Katon et al. [2010] Lin et al. [2012] **Ludman et al. [2013]</p> <p><b>Mehring 2013</b> Mehring et al. [2013]</p> <p><b>Mehring 2014</b> Mehring et al. [2014]</p> <p><b>Meissner 2011</b> Meissner et al. [2011]</p> <p><b>Mortsiefer 2008</b> Mortsiefer et al. [2008] (P)</p> <p><b>Müller-Bühl 2011</b> Müller-Bühl et al. [2011]</p> <p><b>Peters-Klimm 2009</b> Peters-Klimm et al. [2008] *Peters-Klimm et al. [2009] Peters-Klimm et al. [2012]</p>	<p><b>Peters-Klimm 2010</b> Freund et al. [2011a] Peters-Klimm et al. [2007] (P) *Peters-Klimm et al. [2010a]</p> <p><b>Ronel 2011</b> Ronel et al. [2011]</p> <p><b>Rosemann 2007</b> Rosemann et al. [2005] (P) *Rosemann et al. [2007]</p> <p><b>Schencking 2013</b> Schencking et al. [2009] (P) *Schencking et al. [2013]</p> <p><b>Schmidt 2010</b> Schmidt et al. [2010] (P)</p> <p><b>Schmidt 2014</b> Schmidt et al. [2014] (P)</p> <p><b>Schneider 2005</b> **Schneider et al. [2005] *Streitberger et al. [2004]</p> <p><b>Schneider 2006</b> *Schneider et al. [2006] Schneider et al. [2007] Schneider et al. [2008a]</p> <p><b>Siebenhofer 2012</b> Siebenhofer et al. [2012] (P)</p> <p><b>Szecsényi 2011</b> Joos et al. [2005b] (P) *Szecsényi et al. [2011]</p> <p><b>Tinsel 2013</b> Tinsel et al. [2012] (P) *Tinsel et al. [2013]</p> <p><b>Voigt 2011</b> Voigt et al. [2011]</p> <p><b>Vollmar 2007</b> Donath et al. [2010] Holle et al. [2009] (P) Lauterberg et al. [2007] *Menn et al. [2012] **Vollmar et al. [2007b]</p> <p><b>Vollmar 2010</b> Vollmar et al. [2007a] (P) *Vollmar et al. [2010a] Vollmar et al. [2010b]</p> <p><b>Vormfelde 2014</b> Hua et al. [2011] (P) *Vormfelde et al. [2014]</p> <p><b>Zimmermann 2014</b> Zimmermann et al. [2014] (P)</p>
--	---	---

Bei mehr als einer Publikation pro Studie: \* = Hauptpublikation; \*\* = Publikation, welche die Einschlusskriterien in Bezug auf die Autorenschaft erfüllt (falls die Hauptpublikation dies nicht tut); (P) = Studienprotokoll

## **Anhang**

Anhang A: Formular zur Datenextraktion .....	xxx
Anhang B: Schlagwortliste .....	xxxiv
Anhang C: Grafik „Incomplete Outcome Data“ .....	xxxv
Anhang D: Anzahl beteiligter Ärzte/Praxen/Teams .....	xxxvi

## Anhang A: Formular zur Datenextraktion

Autor/Ztschr./Jahr \_\_\_\_\_ Endnote-ID \_\_\_\_\_

**1. THEMA**

1.1	Krankheitsbezogen?	<input type="checkbox"/> ja:			
		<input type="checkbox"/> nein:			
1.2	P				
1.3	I				
1.4	C				
1.5	O				
1.6	Sonstiges				
1.7	ICPC-Code (s6)				<input type="checkbox"/> nicht anwendbar
1.8	ICD-10 Code (s6)				<input type="checkbox"/> nicht anwendbar
1.9	Sprache	<input type="checkbox"/> Englisch	<input type="checkbox"/> Deutsch	<input type="checkbox"/> andere:	

**2. AUTOREN**

2.1	Anzahl Autoren gesamt		2.2	Anzahl Autoren AM			
2.3	Erstautor AM?	<input type="checkbox"/> ja	<input type="checkbox"/> nein	2.4	Letztautor AM?	<input type="checkbox"/> ja	<input type="checkbox"/> nein
2.5	Mitautor Biom.?	<input type="checkbox"/> ja	<input type="checkbox"/> nein	2.6	Mitautor intl.?	<input type="checkbox"/> ja	<input type="checkbox"/> nein
2.7	AM-Abteilungen	1.					
		2.					
		3.					

**3. TEILNEHMENDE ÄRZTE / PRAXEN / TEAMS**

3.1	Zahl (s2) [Praxen/Ärzte]	/	/	<input type="checkbox"/> n/a	
3.2	Art (s3)				<input type="checkbox"/> n/a
3.3	Incentive (s8/9)				<input type="checkbox"/> n/a
3.4	Rekrutierung (s3)				<input type="checkbox"/> n/a
3.5	Arbeitsaufwand (s11/12)				<input type="checkbox"/> n/a
3.6	Datenerhebung (s7)	<input type="checkbox"/> Akte	<input type="checkbox"/> Arzt	<input type="checkbox"/> Patient	<input type="checkbox"/> n/a
		<input type="checkbox"/> sonstiges:			

1

Abbildung 6: Extraktionsformular, Seite 1

Extraktionsformular Kenngrößenreview V6,11.3.2014

**4. PATIENTEN / PROBANDEN**

4.1	Eingeschlossen n (s1)	
4.2	Analysiert n	
4.3	Completer n (s1)	
4.4	Incentive (s10)	O n/a
4.5	Rekrutierung (s4)	O randomisiert
		O individuelle Ansprache
		O konsekutiv
	O Auswahl durch Software	O n/a
	O sonstiges:	
4.6	Arbeitsaufwand (s13)	O n/a

**5. FINANZIERUNG (s23)**

5.1	O DFG	5.2	O BMBF	5.3	O Industrie	5.4	O Eigenmittel
5.5	O Stiftungen	5.6	O anderes:	5.7	O n/a		

**6. STUDIENDESIGN**

6.1	Studienarme	O 2	O 3	O ____	
6.2	Studiendauer / Teilnehm. (s5)			O nur ein Messzeitpunkt	O n/a
6.3	Randomisation	O Patient / Proband	O Arzt	O Praxis	O Clusterrandomisiert
		O andere:			
6.4	falls Cluster-randomisiert (s21/22)	ICC berichtet?	O ja, und zwar:		O nein
		Analyse berücksichtigt Clusterstruktur?	O ja	O nein	
6.5	Verblindung	O Patienten	O Behandler	O Outcomemessung	O keine
		O andere:			
6.6	Intervention Ebenen (s17)	O Arzt	O MFA / Team	O Patient / Proband	
		O andere:			
6.7	Intervention Art (s16)	O Therapie	O Diagnostik	O andere:	
6.8	Fallzahlschätzung (s15)	O ja (geschätzter ICC: _____)			O nein
6.9	ITT-Analyse	O ja	O unklar	O nein	

Abbildung 7: Extraktionsformular, Seite 2

Extraktionsformular Kenngrößenreview V6,11.3.2014

**7. RISK OF BIAS**

	Item	Vote (... risk)	Comment
7.1	A) Random sequence generation	O low O high O unclear	
7.2	B) Allocation concealment	O low O high O unclear	
7.3	C) Blinding of participants and personnel	O low O high O unclear	
7.4	D) Blinding of outcome assessment	O low O high O unclear	
7.5	E) Imcomplete outcome data	O low O high O unclear	
7.6	F) Selective reporting	O low O high O unclear	
7.7	G) Other bias	O low O high O unclear	

**8. ERGEBNISSE**

8.1	Gibt es ein klar vordefiniertes Hauptzielkriterium (HZK)?		
	O ja, nämlich:		O nein
8.2			Wenn nein, gibt es max. drei besonders wichtige Zielkriterien (ZK)?
			O ja, nämlich:      O nein
	Ist genau <u>ein</u> Zeitpunkt für die Erhebung des HZK definiert?		1.
	O ja, und zwar:		2.
	O nein ( => Erhebungszeitpunkte vermerken)		3.
	Erhebungszeitpunkte für die max. drei wichtigen Zielkriterien (ZK) vermerken!		
	Erhebungszeitpunkte	HZK t <sub>1</sub> / t ZK <sub>1</sub>	
	HZK t <sub>2</sub> / t ZK <sub>2</sub>		
	HZK t <sub>3</sub> / t ZK <sub>3</sub>		

Abbildung 8: Extraktionsformular, Seite 3

Extraktionsformular Kenngrößenreview V6,11.3.2014

8.3										
	Authors conclusion					Reviewers conclusion				
	+	(+)	0	(-)	-	+	(+)	0	(-)	-
	Authors conclusion					Reviewers conclusion				
	+	(+)	0	(-)	-	+	(+)	0	(-)	-
	Authors conclusion					Reviewers conclusion				
	+	(+)	0	(-)	-	+	(+)	0	(-)	-

**9. QUERVERWEISE**

9.1	Gibt es Querverweise auf verwandte Studien?	O ja, nämlich	O nein

Abbildung 9: Extraktionsformular, Seite 4

## Anhang B: Schlagwortliste

Stefan Heinmüller

Doktorarbeit Kenngrößenreview

Institut für AM der TUM

### Schlagwortsuche

Item	Deutsch	Englisch
Incentive (3.3, 4.4)	<ul style="list-style-type: none"> <li>• €</li> <li>• Euro</li> <li>• Aufwandsentschädigung</li> </ul>	<ul style="list-style-type: none"> <li>• Incentive</li> <li>• €</li> <li>• Euro</li> <li>• compensation</li> <li>• monetary</li> <li>• [physicians were] reimbursed</li> <li>• received</li> </ul>
Rekrutierung (3.4, 4.5)	<ul style="list-style-type: none"> <li>• Rekrutierung</li> <li>• Konsekutiv</li> <li>• sukzessive</li> </ul>	<ul style="list-style-type: none"> <li>• recruitment</li> <li>• consecutive</li> <li>• successive</li> </ul>
Arbeitsaufwand (3.5, 4.6)	<ul style="list-style-type: none"> <li>• Minuten</li> <li>• Stunden</li> </ul>	<ul style="list-style-type: none"> <li>• minutes</li> <li>• hours</li> </ul>
Eingeschlossen (4.1)	<ul style="list-style-type: none"> <li>• randomisiert</li> </ul>	<ul style="list-style-type: none"> <li>• randomised</li> <li>• randomized</li> </ul>
Finanzierung (5.)	<ul style="list-style-type: none"> <li>• finanziell</li> <li>• finanziert</li> <li>• gefördert</li> </ul>	<ul style="list-style-type: none"> <li>• financial</li> <li>• financed</li> <li>• fund(ed)</li> </ul>
Studiendauer/Teilnehmer (6.2)	<ul style="list-style-type: none"> <li>• Tage</li> <li>• Wochen</li> <li>• Monate</li> </ul>	<ul style="list-style-type: none"> <li>• days</li> <li>• weeks</li> <li>• months</li> </ul>
bei Cluster-RCTs (6.4)	<ul style="list-style-type: none"> <li>• Cluster</li> <li>• ICC</li> <li>• mixed models</li> <li>• generalized estimating equations (GEE)</li> <li>• multi level</li> </ul>	
Verblindung (6.5, 7.3, 7.4)	<ul style="list-style-type: none"> <li>• blind</li> <li>• Unkenntnis</li> </ul>	<ul style="list-style-type: none"> <li>• blind</li> <li>• not informed</li> </ul>
Fallzahlschätzung (6.8)	<ul style="list-style-type: none"> <li>• Fallzahl</li> </ul>	<ul style="list-style-type: none"> <li>• sample size</li> <li>• „we aimed for X patients“</li> <li>• X patients were needed</li> </ul>
ITT-Analyse (6.9)	<ul style="list-style-type: none"> <li>• intention-to-treat (ITT)</li> </ul>	
Sequence generation (7.1)	<ul style="list-style-type: none"> <li>• Computer</li> <li>• Los</li> </ul>	<ul style="list-style-type: none"> <li>• Computer-based</li> <li>• lot</li> </ul>
Allocation concealment (7.2)	<ul style="list-style-type: none"> <li>• zentral</li> </ul>	<ul style="list-style-type: none"> <li>• central allocation</li> </ul>
Hauptzielkriterium (8.1)	<ul style="list-style-type: none"> <li>• (Haupt-)Zielkriterium</li> </ul>	<ul style="list-style-type: none"> <li>• main/primary outcome/endpoint</li> </ul>
verwandte Studien (9.1)	<ul style="list-style-type: none"> <li>• vorhergehende (Studie)</li> </ul>	<ul style="list-style-type: none"> <li>• previous (study)</li> </ul>

Abbildung 10: Liste für Schlagwortsuche

## Anhang C: Grafik „Incomplete Outcome Data“

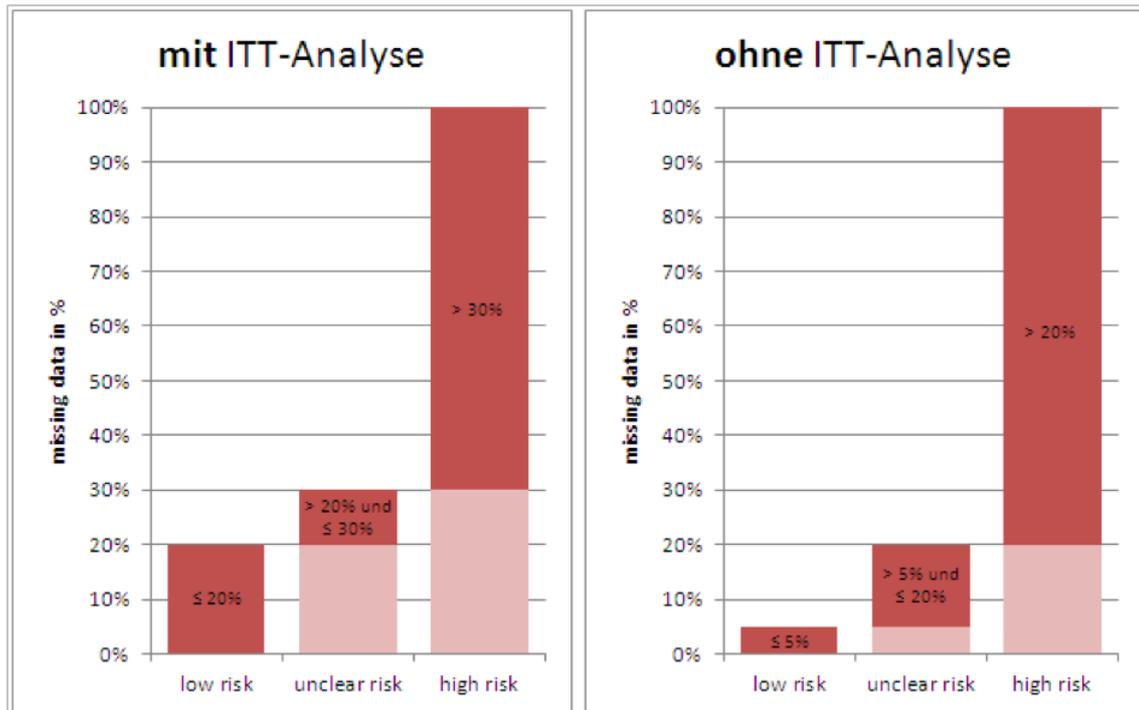


Abbildung 11: Beurteilung der Verzerrungsrisiken anhand von ITT-Analyse und „fehlenden Daten“

## **Anhang D: Anzahl beteiligter Ärzte/Praxen/Teams**

Wie unter 3.2.7.2 beschrieben, sollte für die 39 in Praxen (Arztpraxen und Praxen anderer Heilberufe) stattfindenden multizentrischen Studien (Tabelle 25) die Anzahl der zum Zeitpunkt  $t_0$  (initial) an der Studie beteiligten Praxen (bzw. der randomisierten Praxen im Falle einer Clusterrandomisierung), der zu  $t_0$  an der Studie beteiligten Ärzte<sup>1</sup>, sowie die Anzahl der tatsächlich beteiligten Praxen und Ärzte zum Zeitpunkt  $t_1$  (dem Zeitpunkt der ersten Datenerhebung) ermittelt werden. Die Unterscheidung zwischen Praxen und Ärzten war nötig, da deren Anzahl voneinander abweichen konnte, wenn sich unter den beteiligten Praxen Gemeinschaftspraxen mit mehreren Ärzten befanden. Die Gewinnung dieser Informationen gestaltete sich jedoch häufig schwierig, da diese in vielen Publikationen nur unvollständig genannt wurden. So wurden „Ärzte“ und „Praxen“ von vielen Autoren synonym verwendet, nicht selten war nur die Anzahl von Ärzten *oder* Praxen genannt oder die Informationen waren nur für einen Zeitpunkt verfügbar, sodass die Anzahl der Praxen bzw. Ärzte, die ihre Studienteilnahme vorzeitig beendet hatten, nicht eindeutig bestimmt werden konnte.

Vor dem Hintergrund dieser Problematik entschied sich der Autor dieser Arbeit für folgendes Vorgehen: für all diejenigen Studien, bei denen sowohl die Anzahl der beteiligten Praxen als auch die Anzahl der beteiligten Ärzte bekannt waren, wurde der entsprechende Steigerungsfaktor berechnet. Aus diesen Steigerungsfaktoren wurde der Mittelwert gebildet und dieser gerundet. Auf diesem Weg konnte folgende Formel erstellt werden, die eine annäherungsweise, durchschnittliche Ermittlung der Anzahl von Ärzten/Praxen aus dem jeweils anderen Wert ermöglichte:

$$\text{Anzahl beteiligter Ärzte} = \text{Anzahl beteiligter Praxen} * 1,065$$

Für all diejenigen Studien, bei denen sich nur eine dieser beiden Zahlen (Ärzte oder Praxen) erheben ließ, wurde die jeweils andere Zahl mithilfe der genannten Formel berechnet.

Die Anzahlen beteiligter Ärzte/Praxen zu den verschiedenen Erhebungszeitpunkten können Tabelle 33 entnommen werden.

---

<sup>1</sup> auch wenn sich hierbei in jeweils einer Studie um Zahnarzt- bzw. physiotherapeutische Praxen handelte, wird der Einfachheit weiterhin von „Ärzten“ gesprochen

Tabelle 33: Anzahl der beteiligten Praxen/Ärzte zu Studienbeginn ( $t_0$ ) und zum Zeitpunkt der ersten Datenerhebung ( $t_1$ )\*

Anzahl beteiligter Praxen/Ärzte	RCTs mit Ergebnissen	RCTs ohne Ergebnisse (Studienprotokolle)	Gesamt	Cluster-RCTs	„klassische“ RCTs
initial ( $t_0$ ) beteiligte Praxen	39,5 (10–286) n = 24	48 (20–188) n = 15	46 (10–286) n = 39	65 (13–286) n = 28	20 (10–45) n = 11
initial ( $t_0$ ) beteiligte Ärzte	41 (10–303) n = 24	52 (22–201) n = 15	49 (10–303) n = 39	69,5 (22–303) n = 28	22 (10–48) n = 11
zum Zeitpunkt der ersten Datenerhebung ( $t_1$ ) beteiligte Praxen	38,5 (10–167) n = 24	-	38,5 (10–167) n = 24	46,5 (22–167) n = 16	17,5 (10–29) n = 8
zum Zeitpunkt der ersten Datenerhebung ( $t_1$ ) beteiligte Ärzte	40 (10–178) n = 24	-	40 (10–178) n = 24	56 (24–178) n = 16	19,5 (10–31) n = 8

\* dargestellt als Median (Minimum–Maximum)

## Publikationsverzeichnis

Die Ergebnisse der vorliegenden Übersichtsarbeit wurden in zwei Artikeln veröffentlicht:

Heinmüller, S., Schneider, A., Linde, K.

*Quantity, topics, methods and findings of randomised controlled trials published by German university departments of general practice – systematic review.*

Trials. 17 (2016a) 211

Heinmüller, S., Schneider, A., Linde, K.

*Randomisierte Studien der deutschen universitären Allgemeinmedizin – Ergebnisse einer systematischen Übersichtsarbeit.*

Z. Allgemeinmed. 92 (2016b) 259-264

## Danksagung

Besonders danken möchte ich meinem Doktorvater, Herrn Professor Linde, der den gesamten Entstehungsprozess der vorliegenden Doktorarbeit engmaschig begleitete. Stets skizzierte er den jeweils nächsten Arbeitsschritt strukturiert und lieferte auf diesem Weg sehr hilfreiche Richtungsvorgaben. Bei Problemen während des Arbeitsprozesses war Herr Professor Linde immer kurzfristig verfügbar und bot wertvolle Hilfestellung. Die jedem Arbeitsschritt nachfolgende Rückmeldung erfolgte zeitnah und beinhaltete durchwegs konstruktive Verbesserungsvorschläge. Herr Professor Linde war maßgeblich an der Erstellung der beiden mit dieser Arbeit assoziierten Publikationen beteiligt und führte die Metaanalyse (3.4.2) durch, deren Ergebnisse ich freundlicherweise in der vorliegenden Arbeit verwenden durfte.

Mein Dank gilt weiterhin Herrn Professor Schneider, der mir für einen großen Teil der Forschungsarbeit einen Arbeitsplatz im Institut für Allgemeinmedizin der TU München zur Verfügung stellte und wertvolle Anregungen für meine Doktorarbeit gab.

Bis zu seinem Ausscheiden aus dem Institut für Allgemeinmedizin der TU München wurde ich zudem von Herrn Michael Barth betreut, der mir bei der Literatursuche mit Scopus und der Erstellung des Extraktionsformulars mit Rat und Tat zur Seite stand, wofür ich ihm herzlich danken möchte.

Bei dem restlichen Team des Instituts für Allgemeinmedizin der TU München möchte ich mich ebenfalls bedanken. Durch ein sehr angenehmes Arbeitsklima haben auch die Kollegen zur Entstehung dieser Dissertation beigetragen.

Danken möchte ich weiterhin der Stiftung der Deutschen Wirtschaft, die mich durch das Studienförderwerk Klaus Murmann über große Teil meines Studiums hinweg unterstützte und mir für die Erstellung der Dissertation zwei weitere Fördersemester gewährte.

Ein besonderer Dank gilt meiner Schwester Johanna Heinmüller, die mir geduldig mit der Textverarbeitung half, sowie meinen Eltern, Dr. Mechthild und Christian Heinmüller, für das Korrekturlesen dieser Arbeit und die stete Unterstützung über die gesamte Dauer meines Studiums und der Dissertation.