Aus dem Institut für Arbeits-, Sozial- und Umweltmedizin
(Prof. Dr. med. Ernst Hallier)
der Medizinischen Fakultät der Universität Göttingen

Untersuchung arbeitsbezogener Endpunkte in randomisierten, kontrollierten Studien zur Behandlung chronischer Schmerzerkrankungen

INAUGURAL-DISSERTATION

zur Erlangung des Doktorgrades für Zahnheilkunde

der Medizinischen Fakultät der

Georg-August-Universität zu Göttingen

vorgelegt von

Ingmar Wolf

aus Darmstadt

Göttingen 2015

Diese Arbeit wurde angefertigt unter Anleitung der gleichberechtigten Betreuer Prof. Dr. Sebastian Straube (Division of Preventive Medicine, University of Alberta; ehemals Institut für Arbeits-, Sozial- und Umweltmedizin, Universitätsmedizin Göttingen) und Prof. Dr. Tim Friede (Institut für Medizinische Statistik, Universitätsmedizin Göttingen) im Zeitraum von Juli 2011 bis Januar 2015.

Dekan:	Prof. Dr. rer. nat. H. K. Kroemer
I. Berichterstatter:	Prof. Dr. Sebastian Straube
II. Berichterstatter:	Prof. Dr. Frank Petzke
Tag der mündlichen Prüfung:	10.08.2016

Inhaltsverzeichnis

Abkürzungsverzeichnis.....7

1. Einleitung.....8

1.1 Zielsetzung.....8

1.2 Was sind chronische Schmerzen?.....8

1.3 Wer ist von chronischen Schmerzen betroffen und welche Auswirkungen haben sie auf die Arbeitswelt?.....9

1.4 Erkrankungen mit chronischen Schmerzen.....10

1.4.1 Chronische Rückenschmerzen.....10

1.4.2 Fibromyalgie.....11

1.4.3 Rheumatoide Arthritis.....12

1.4.4 Sonstige relevante Erkrankungen.....14

1.5 Erfassung von arbeitsbezogenen Endpunkten in klinischen Studien.....14

1.6 Die Datenbank für die Literaturrecherche.....16

2. Material und Methoden.....18

2.1 Die systematische Literaturrecherche.....18

2.2 Bewertung der Studienqualität.....21

2.3 Datenextraktion.....22

2.4 Statistische Verfahren.....24

2.4.1 Meta-Analysen und Heterogenität.....24

2.4.2 Modelle der Meta-Analysen.....25

2.4.3 Funnel-Plots.....26

2.4.4 Standardisierung der Studienergebnisse.....28

2.4.5 Präsentation der Meta-Analysen.....29

2.4.6 Qualitative Endpunkte.....30

2.4.7 Trends und Regressionen.....31

2.4.8 „Votecount“.....32

3. Ergebnisse.....33

3.1 Studienübersicht.....33

3.2 Untersuchung von arbeitsbezogenen Endpunkten in der Literatur 39

3.3 Wie wurde über arbeitsbezogene Endpunkte berichtet?.....42

3.4 Studienergebnisse.....45

3.4.1 Gesamtüberblick.....45

3.4.1.1 Meta-Analyse der Durchschnittsdifferenzen aller arbeitsbezogenen Endpunkte.....47

3.4.1.2 Heterogenitäts- und Sensitivitätsanalyse.....48

3.4.1.3 Meta-Analyse aller arbeitsbezogenen Endpunkte (Daten zu Studienende).....51

3.4.2 Meta-Analyse zu „Brief Pain Inventory: Interference with normal work“.....52

3.4.3 Meta-Analyse zu „Interference with work“ (BPI/FIQ).....53

3.4.3.1 Vergleich der Ergebnisse auf der BPI/FIQ-Skala und der standardisierten Ergebnisse.....54

3.4.4 Meta-Analyse zu Arbeitsfehlzeiten.....55

3.4.4.1 Meta-Analyse zu Arbeitsfehlzeiten berichtet als Durchschnittsdifferenzen.....55

3.4.4.2 Meta-Analyse zu Arbeitsfehlzeiten berichtet zu Studienende.....56

3.4.4.3 Vergleich der Ergebnisse aus den Abschnitten 3.4.4.1 und 3.4.4.2.....57

3.4.5 Arbeitsproduktivität.....57

3.4.5.1 Meta-Analyse zu erfasster Arbeitsproduktivität.....58

3.4.5.2 Meta-Analyse zu erfasster Arbeitsproduktivität nur der Visual-Analog-Scale-Studien.....58

3.4.5.3 Vergleich der Ergebnisse auf der VAS-Skala und der standardisierten Ergebnisse.....59

3.4.6 Meta-Analyse zu „Brief Pain Inventory: Interference with normal work“ bei chronischen Rückenschmerzen.....60

3.4.7 Arbeitsfähigkeit.....60

3.4.7.1 Meta-Analyse zu der Vermeidung von Arbeitsunfähigkeit.....61

3.4.7.2 Meta-Analyse zu der Wiederherstellung von Arbeitsfähigkeit.....62

3.4.7.3 Vergleich der Ergebnisse aus den Abschnitten 3.4.7.1 und 3.4.7.2.....63

3.4.7.4 Meta-Analyse zu der Arbeitsfähigkeit zu Studienende 63

3.5 Schmerzbezogene Endpunkte.....64

3.5.1 Meta-Analysen der schmerzbezogenen Endpunkte.....65

3.5.1.1 Meta-Analyse der Durchschnittsdifferenzen zu schmerzbezogenen Endpunkten.....66

3.5.1.2 Meta-Analyse der Enddaten zu schmerzbezogenen Endpunkten....67
3.5.1.3 Heterogenitäts- und Sensitivitätsanalyse....67
3.5.2 Vergleich arbeitsbezogener Endpunkte mit Schmerzendpunkten....68
3.5.2.1 Regression zu schmerz- und arbeitsbezogenen Endpunkten des „Brief Pain Inventory“....69
3.5.2.2 Regression zu allen schmerz- und arbeitsbezogenen Endpunkten....70
3.6 Ansprechen auf Behandlungen....72
3.6.1 Meta-Analysen der Ansprechraten für schmerzbezogene Endpunkte....72
3.6.1.1 Meta-Analysen der 30%-Ansprechraten....74
3.6.1.2 Meta-Analysen der 50%-Ansprechraten....75
3.6.1.3 Heterogenitäts- und Sensitivitätsanalysen....76
3.6.2 Meta-Analysen der ACR-Ansprechraten....77
3.6.2.1 Meta-Analysen der 20%-ACR-Ansprechraten....78
3.6.2.2 Meta-Analysen der 50%-ACR-Ansprechraten....79
3.6.2.3 Meta-Analysen der 70%-ACR-Ansprechraten....80
3.6.2.4 Heterogenitäts- und Sensitivitätsanalysen....80
3.7 Ansprechraten und arbeitsbezogene Endpunkte....85
3.7.1 Regressionen der schmerzbezogenen 30%-Ansprechraten mit den standardisierten, arbeitsbezogenen Endpunkten. .86
3.7.2 Regressionen der schmerzbezogenen 50%-Ansprechraten mit den standardisierten, arbeitsbezogenen Endpunkten. .87
3.7.3 Heterogenitäts- und Sensitivitätsanalysen zu den Abschnitten 3.7.1 und 3.7.2....88

4. Diskussion....89
4.1 Einschränkungen der verwendeten Modelle und Verfahren....89
4.1.1 Vollständigkeit der Literaturrecherche....89
4.1.2 Studienqualität....89
4.1.3 Datenextraktion aus Interquartilangaben....94
4.2 Heterogenität....96
4.2.1 Kontinuierliche Endpunkte....96
4.2.1.1 Standardisierung der Studienergebnisse....96
4.2.1.2 Transformierung der SMD in klinische Endpunkte. .97
4.2.1.3 Transformierung der SMD in den odds ratio....98
4.2.2 Ergebnisinterpretation....100
4.2.3 Ansprechraten....101

4.3 Arbeitsbezogene Endpunkte in der Literatur.....102

4.4 Erfassung von arbeitsbezogenen Endpunkten in zukünftigen Arbeiten.....104

4.5 Ansprechraten für arbeitsbezogene Endpunkte und die Berichterstattung des Behandlungseffektes.....108

5. Zusammenfassung.....110

Literaturverzeichnis.....113

Anhang.....127

Anhang 1.....127

Anhang 2.....152

Anhang 3.....154

Anhang 4.....157

Publikationen im Zusammenhang mit dieser Dissertation.....159

Abkürzungsverzeichnis

ACR = American College of Rheumatology
ANOVA = *analysis of variance*= Varianzanalyse
BMI = Body Mass Index
BOCF = *baseline observation carried forward*
BPI = Brief Pain Inventory
DASH = Disabilities of the Arm, Shoulder and Hand Questionnaire
DMARD = *disease modifying antirheumatic drugs*
FE = *fixed effect model*
FIQ = Fibromyalgia Impact Questionnaire
ITT = Intention-To-Treat-Analyse
KI = Konfidenzintervall
LOCF = *last observation carried forward*
LWDE = *lost workday equivalents*
MAF = Multidimensional Assessment of Fatigue
MeSH = Medical-Subject-Headings-Datenbank
MMRM = *mixed model for repeated measurements*
MTX = Methotrexat
NSAR/NSAP = nichtsteroidalen Antirheumatika/Antiphlogistika
OQS = Oxford Quality Scale
OR = *odds ratio*
RD = *risk difference* = Wahrscheinlichkeitsdifferenz
RE = *random effects model*
RR = *risk ratio* = Wahrscheinlichkeitsquotient
SDS = Sheehan Disability Scale
SF-36 = Short Form 36 Health Survey Questionnaire
SMD = standardisierte Mittelwertsdifferenz
TENS = *transcutaneous electrical nerve stimulation* = Reizstromtherapie
TNF = Tumornekrosefaktor
VAS = visuelle Analogskala
WHO = World Health Organization
WIS = Work Instability Scale
WPAI = Work Productivity and Activity Impairment Questionnaire
WPS-RA = Rheumatoid Arthritis Specific Work Productivity Survey

1. Einleitung

1.1 Zielsetzung

Ziel der vorliegenden Dissertation ist es, einen möglichst vollständigen Überblick über die Erfassung von arbeitsbezogenen Endpunkten[1] und deren Entwicklung bei Behandlung in klinischen, randomisierten Studien über chronische Schmerzerkrankungen und schmerzhafte, rheumatische Erkrankungen zu geben. Dazu wurde eine systematische Literaturrecherche durchgeführt und anschließend ein systematisches Review erstellt. Die in den gefundenen und ausgewerteten Studien berichteten Daten wurden, wenn möglich, in Meta-Analysen zusammengefasst.

1.2 Was sind chronische Schmerzen?

Nach der "Internationalen statistischen Klassifikation der Krankheiten und verwandter Gesundheitsprobleme" in der aktuellen, deutschen Ausgabe von 2013 (ICD-10-GM) werden anhaltende Schmerzstörungen in somatoforme Störungen und chronische Schmerzstörungen mit somatischen und psychischen Faktoren unterteilt. Während den somatoformen Erkrankungen eine unbekannte oder psychosoziale Ursache zu Grunde liegt, stützten sich die letzteren Störungen zumindest zu Erkrankungsbeginn auf eine nachweisbare pathologische Veränderung. Als Mindestdauer der Schmerzen gibt die WHO sechs Monate an.
In der Literatur werden chronische Schmerzen weitaus weniger scharf umgrenzt. So sind andauernde Schmerzen von drei bis sechs Monaten als Abgrenzungsmerkmal in klinischen Studien weit verbreitet. Auch variable Zeiträume, welche die als normal angenommene Zeit bis zur Heilung

1 Endpunkt einer Studie: Zu erfassende Untersuchungsergebnisse wie beispielsweise empfundener Schmerz oder Arbeitszeitverluste

überschreiten, finden bei der Definition von chronischen Schmerzen Anwendung (Debono et al. 2013).

1.3 Wer ist von chronischen Schmerzen betroffen und welche Auswirkungen haben sie auf die Arbeitswelt?

Die groß angelegte, von Breivik et al. 2006 durchgeführte Gesundheitsumfrage gewährte einen Einblick, inwieweit die europäische Bevölkerung von chronischen Schmerzen betroffen ist, und wie sich diese auf einzelne Lebensbereiche auswirken. Danach leidet fast jeder fünfte erwachsene Europäer unter chronischen Schmerzen mit einer Symptomdauer von mindestens 6 Monaten. Von diesen sind immerhin 21% über einen Zeitraum von 20 Jahren oder länger betroffen. Bei fast der Hälfte der Befragten sei der Rücken und bei ca. 40% die Gelenke betroffen. Als spezifische Krankheitsbilder nannten 42% der Patienten Osteoarthritis oder rheumatoide Arthritis, und 15% berichteten über chronische Schmerzen aufgrund von degenerierten beziehungsweise prolabierten Bandscheiben. Die Auswirkungen auf das Arbeitsleben betreffend berichteten 19% der nicht verrenteten Patienten, dass sie wegen ihrer Erkrankung arbeitslos seien, 16% mussten ihre Arbeitsschwerpunkte verschieben und 13% arbeiteten in einem anderen Beruf. Durchschnittlich versäumten die arbeitenden Befragten 7,8 Arbeitstage während des vergangen halben Jahres aufgrund ihrer Schmerzen. Maniadakis und Gray schätzten im Jahr 2000, dass in Großbritannien jährlich umgerechnet ca. 10,6 Milliarden Euro wegen ausgefallener Arbeitszeit aufgrund von chronischen Rückenschmerzen verloren gegangen wären.
Die Verbesserung der Arbeitsfähigkeit durch Behandlungen ist daher nicht nur für das Allgemeinbefinden des Individuums relevant, sondern ist auch für die gesamte Volkswirtschaft von Bedeutung.

1.4 Erkrankungen mit chronischen Schmerzen

1.4.1 Chronische Rückenschmerzen

Etwa 50% der Europäer, die unter chronischen Schmerzen leiden, beschreiben diese als Rückenschmerzen (Breivik et al. 2006). Chronische Rückenschmerzen beziehen sich meist auf Schmerzen unter der Rippenregion und werden in der deutschen Sprache oft als Kreuzschmerzen oder Lumbalgie bezeichnet, während in der englischsprachigen Literatur oft von *low back pain* die Rede ist. Die meisten Rückenschmerzen erscheinen unspezifisch, also ohne erkennbare Ursache (Manek und MacGregor 2005). Bei spezifischen Rückenschmerzen sind klare Befunde auszumachen, welche die Schmerzen verursachen könnten. Dazu zählen Bandscheibenvorfälle, Frakturen der Wirbel, Tumore und Infektionen. Die Prävalenz dieser Befunde bei Rückenschmerzen ist gering und überschreitet einzeln betrachtet nicht einmal 5% aller Diagnosen bei Rückenschmerzen (van den Bosch et al. 2004).

Die genauen Ursachen sowie die Entwicklung einer Chronifizierung sind nicht genau bekannt. Es werden jedoch Faktoren wie Übergewicht, psychologische Einflüsse und übermäßige, einseitige Tätigkeiten vermutet (Manek und MacGregor 2005). Mit zunehmendem Alter treten Bandscheibendegeneration und Osteophytenbildung häufiger auf, die ebenfalls mit Rückenschmerzen assoziiert sind (Pye et al. 2004).

Zur symptomatischen Behandlung kommen meist nichtsteroidale Antirheumatika oder schwache Opioide wie Tramadol zum Einsatz, die Schmerzen mildern können (Roelofs et al. 2008 und Chaparro et al. 2013). Außerdem scheinen Massagen das Potenzial für eine Verbesserung von chronischen Rückenschmerzen zu haben, während Chirotherapie und Reizstromtherapie (TENS) anscheinend keine Verbesserung zu erreichen vermögen (Furlan et al. 2008, Rubinstein et al. 2012, Khadilkar et al. 2008). Weiterhin scheinen bei chronischen Rückenschmerzen auf Bewegung basierte Therapien eine leichte Schmerzreduzierung zu bewirken (Hayden et al. 2005).

1.4.2 Fibromyalgie

Fibromyalgie ist ein über den Körper generalisiertes Schmerzsyndrom, bei dem die genauen Ursachen bisher noch im Dunkeln liegen.

Für das Erfassen einer möglichst homogenen Patientengruppe hat sich die Definition des American College of Rheumatology von Wolfe et al. 1990 etabliert. Darin wird der Körper in ein Koordinatensystem mithilfe einer horizontalen Linie im Bereich der Taille sowie einer Linie im vertikalen Median unterteilt. Alle Quadranten sollten einen schmerzhaften Bereich aufweisen. Weiterhin definierte die Arbeitsgruppe 18 Stellen am Körper, die sogenannten Tenderpoints, von denen bei einer digitalen Druckaufwendung von etwa 4 kg mindestens 11 der 18 Punkte für eine Fibromyalgiediagnose schmerzhaft reagieren sollten.

Bei der Prävalenz werden Werte zwischen 0,66% bis 3,3% dokumentiert. Einheitlich wird von deutlich mehr betroffenen Frauen als Männern berichtet (Neumann und Buskila 2003). Wie Nimnuan et al. 2001 zeigen konnten, tritt das Geschehen mit weiteren unscharfen Syndromen wie Reizdarm, Spannungskopfschmerzen, Verdauungsstörungen, Hyperventilation und nicht kardial begründeten Brustschmerzen auf. Ebenso vielfältig sind auch die möglichen Ätiologien der vermutlich multifaktoriellen Erkrankung. Als eine der Hauptursachen werden fehlerhafte Vorgänge im zentralen Nervensystem angesehen. So konnten Staud et al. 2001 eine erhöhte Schmerzreaktion nach vorheriger neuronaler Reizungen bei Fibromyalgie-Patienten im Vergleich zu gesunden Probanden nachweisen. Dies lässt auf gestörte Schmerzverstärkungsmechanismen im dorsalen Horn des Rückenmarkes schließen. Daneben wurden Zusammenhänge mit endokrinen Veränderungen im Serotonin-, Noradrenalin-, Dopamin- oder Substanz-P-Haushalt, mit Schlafstörungen, mit genetischen Konstellationen, mit Autoimmunreaktionen, mit psychischen Störungen und mit Infektionen gefunden (Bellato et al. 2012).

Die Therapieversuche reichen von Bewegungstherapie, Akupunktur, psychologischen Interventionen bis hin zu medikamentösen Behandlungen. In

verschiedenen Reviews untersuchten Autoren für die Cochrane Collaboration den Evidenzstatus unterschiedlicher Behandlungen. Demnach hat Akupunktur vermutlich keinen Effekt auf die Schmerzen (Deare et al. 2013). Tort et al. konnten 2012 in ihrem Review moderate Evidenz für einen positiven Effekt von Monoaminooxidase-Hemmern wie Moclobemid feststellen. Ebenso scheinen Seretonin-Noradrenalin-Wiederaufnahme-Inhibitoren wie Duloxetin und Milnacipran einen leichten Effekt bei der Schmerzbehandlung zu haben (Häuser et al. 2013). Weiterhin erhöht betreutes Aerobic-Training das allgemeine Wohlbefinden und den physischen Allgemeinzustand von Fibromyalgiepatienten, kann aber vermutlich nicht die Schmerzen lindern (Busch et al. 2007). Amitriptylin, ein Mittel der ersten Wahl bei neuropathischen Schmerzen, kann dagegen nur eine eingeschränkte Evidenz vorweisen (Moore et al. 2012a). Effektive Behandlungen konnten mit Pregabalin beobachtet werden (Moore et al. 2009). Weiterhin konnte gezeigt werden, dass transkranielle Gleichstromstimulation helfen kann, das Schmerzempfinden von Fibromyalgiepatienten sowie von Patienten mit chronischen Rückenschmerzen positiv und nachhaltig zu modulieren (Antal et al. 2010).

1.4.3 Rheumatoide Arthritis

Rheumatoide Arthritis ist eine durch Synovitis geprägte degenerative und progrediente Erkrankung. Nach den Kriterien der „American Rheumatism Association“ liegt sie vor, wenn Morgensteifigkeit der Gelenke für mindestens eine Stunde besteht, Gelenke der Hände oder das Handgelenk selbst geschwollen sind, insgesamt 3 oder mehr geschwollene Gelenke vorliegen und die arthritischen Schwellungen auf beiden Körperhälften symmetrisch auftreten. Diese Symptome müssen länger als 6 Wochen beobachtet worden sein. Weiterhin sind Rheumaknoten, Rheumafaktoren und Gelenkerosionen oder Knochendichteverlust in der Hand sichere Anzeichen für eine rheumatoide Arthritis. Von diesen Anzeichen sollten mindestens 4 nachgewiesen werden. (Arnett et al. 1988). Als Rheumaknoten werden Bindegewebsveränderungen im

Subkutangewebe bezeichnet. Diese deuten meist einen schweren Verlauf der Krankheit an. Rheumafaktoren sind Autoimmunanitkörper, welche an verschiedene andere Immunglobuline binden können. Weitere serologische Bedeutung haben Antikernfaktoren und Immunkomplexe des Komplement-systems. Der genaue Einfluss dieser Faktoren sowie der detaillierte pathogenetische Ablauf der Erkrankung sind noch unbekannt (Hettenkofer et al. 2003). Es wird vermutet, dass genetische Prädisposition, vor allem assoziert mit HLA-Genen (Nepom et al. 1989, Muazzam et al. 2013), aber auch mit anderen Genloci wie ARL15 (Negi et al. 2013), eine Autoimmunreaktion begünstigt, welche die Destruktionen an Bindegewebe, Organen und Gelenken bedingt.

Nach Symmons et al. 2002 liegt die Prävalenz in der britischen Bevölkerung bei etwa 1,16% bei Frauen und etwa 0,44% bei Männern.
Im Verlauf der Krankheit stellen sich arthritische Schmerzen sowie Funktionseinschränkungen durch Degeneration der Gelenke und des Sehnengleitgewebes ein. Später können praktisch alle Organe, Blutgefäße und auch Nerven befallen sein (Hettenkofer et al. 2003).
Zur Behandlung kommen vor allem immunmodulierende Medikamente sowie Bewegungstherapien zum Einsatz. Im Rahmen verschiedener systematischer Reviews für die Cochrane Collaboration wurden mehrere Medikamente auf ihre Wirkung hin untersucht. Als Basistherapie wird oft eine Kombination aus Glukokortikoiden mit *disease modifying antirheumatic drugs* (DMARDs) gegeben. In ihrer Meta-Analyse von 2007 berichteten Kirwan et al., dass Glukokortikoide das Fortschreiten von Gelenkerosionen deutlich reduzieren. Als eines der häufigsten DMARDs wird erfolgreich das Zytostatikum Methotrexat (MTX) eingesetzt (Suarez-Almazor et al. 1998). Weiterhin werden der Entzündungshemmer Leflunomid (Osiri et al. 2003) und der Immunsuppressor Cyclosporin (Wells et al. 1998) sowie weitere Stoffgruppen verabreicht. In den letzten Jahren kamen Biologika, von gentechnisch veränderten Bakterien hergestellte Medikamente, auf den Markt. So konnten mit dem T-Zell-Repressor Abatacept, dem Interleukin-1-Antagonisten Anakinra, dem B-Zell-Depletor Rituximab und den TNF-Blockern Etanercept, Infliximab und Adalimumab gute Ergebnisse erzielt werden (Singh et al. 2009).

1.4.4 Sonstige relevante Erkrankungen

Weitere relevante entzündliche rheumatische Erkrankungen nach Hettenkofer et al. 2003 sind:

- Spondylitis ankylosans, welche sich durch chronische Rückenschmerzen und eine Versteifung der Wirbelgelenke auszeichnet.
- Die reaktive Arthritis, welche nach einer durchgemachten Infektion in verschiedenen Ausprägungen auftritt.
- Psorisasarthritis, ein Syndrom, bei dem Arthriden mit Psorisas-Läsionen kombiniert auftreten.

Dagegen werden degenerative, primär nicht entzündliche, Gelenkerkrankungen abgegrenzt. Die primäre Arthrose (im Englischen *osteoarthritis*) tritt ohne spezifischen Grund auf. Sind Deformationen, Fehlbelastungen oder andere negative Einflüsse nachweisbar, so wird von einer sekundären Arthrose gesprochen. Im Alter nehmen Knorpel- und Knochendegenerationen durch Verschleiß deutlich zu. So sind Personen mit 70 Jahren zu 90% von Gelenkschäden betroffen. Von diesen berichten jedoch nur etwa 50% von Beschwerden (Miehlke und Schmidt 2003).

1.5 Erfassung von arbeitsbezogenen Endpunkten in klinischen Studien

Arbeitsbezogene Endpunkte werden zur Erleichterung der praktischen Durchführung in prospektiven Studiendurchführungen vornehmlich mithilfe von Patientenbefragungen untersucht. Die Datenerfassung ist als Grundstein jeder statistischen Auswertung von zentraler Bedeutung. Daher ist die Auswahl beziehungsweise die Gestaltung der hierzu eingesetzten Fragebögen sehr wichtig (Olsen 1998). Die betreffenden Abschnitte der Fragebögen sollten den zu untersuchenden Endpunkt unmissverständlich und ohne Interpretationsspielraum vom Probanden erfragen. Weiterhin sollten die Formulierungen der Fragen

neutral und ergebnisoffen gestaltet sein, so dass dem Patienten keine Antwort suggeriert wird. Eine einheitliche, überschneidungsfreie Formatierung der Antwort, beispielsweise auf vorgegebenen Skalen, erleichtert zusätzlich die spätere Auswertung. Abschließend sollten Pilotstudien durchgeführt werden, wodurch eventuell auftretende Schwächen des Fragebogens aufgedeckt und bei Bedarf verbessert werden können (Stone 1993). Bei der Validierung werden in prospektiven klinischen Studien die neuen Messinstrumente mit bereits gebräuchlichen Instrumenten verglichen. Dabei soll festgestellt werden, wie groß die Korrelation zwischen den verschiedenen Datenerhebungen ist und wie stabil sich die Ergebnisse bei einem wiederholten Test darstellen. Ein so validiertes Instrument sollte verlässlich den zu untersuchenden Endpunkt, unabhängig von den durchführenden Personen, möglichst objektiv dokumentieren.
Verbreitete und validierte Fragebögen sind:

- Das „Brief Pain Inventory“ (BPI) wird häufig bei chronischen Schmerzen im Zusammenhang mit malignen und non-malignen Erkrankungen verwendet. Mithilfe des BPI werden selbst empfundene Schmerzen sowie unter anderem arbeitsrelevante Einschränkungen des täglichen Lebens dokumentiert. Die Übersetzungen dieses Instrumentes in verschiedene Sprachen wurden ausführlich in verschiedenen Studien validiert (Radbruch et al. 1999).
- Das „Fibromyalgia Impact Questionnaire“ (FIQ) hat sich für die Befragung von Fibromyalgiepatienten etabliert. Hierbei interessieren vornehmlich die Einschränkungen, welche die Patienten durch die Erkrankung erfahren. Wie das zuvor erwähnte BPI wurden auch die verschiedenen Übersetzungen des FIQ eingehend analysiert und validiert (Offenbaecher et al. 2000).
- Das „Short Form 36 Health Survey Questionnaire“ ist ein breit gefächerter Fragenkatalog, der in vielen epidemiologischen Befragungen zum Einsatz kommt. Mit diesem validierten Instrument wird der allgemeine Gesundheitsstatus eines Patienten erfasst (Brazier et al. 1992).

- Das „Work Productivity and Activity Impairment Questionnaire“ (WPAI) konzentriert sich dagegen gänzlich auf die Erfassung arbeitsrelevanter Endpunkte (Zhang et al. 2010).
- Die „Rheumatoid Arthritis Specific Work Productivity Survey“ (WPS-RA) erfasst zusätzlich zu den arbeitsbezogenen Leistungen die Art der ausgeübten Tätigkeit (Osterhaus et al. 2009).
- Die „Work Instability Scale“ (WIS) soll die Möglichkeit eines krankheitsbedingten Erwerbsausfalles genauer erfassen (Gilworth et al. 2009).

1.6 Die Datenbank für die Literaturrecherche

Die systematische Literaturrecherche wurde unter Zuhilfenahme des MEDLINE®/PubMED® - Systems durchgeführt. MEDLINE® wird von der „US National Library of Medicine“ gepflegt und hat sich mit über 40 Jahren Betriebszeit als die bekannteste und am häufigsten verwendete, öffentlich zugängliche Literaturdatenbank in den Bereichen Medizin, Biochemie und Biologie etabliert. Das ursprüngliche System unterstützte 239 wissenschaftliche Zeitschriften und 25 gleichzeitige Suchanfragen (Lindberg 2000). Inzwischen werden 5640 Journals erfasst und 2,5 Milliarden Suchanfragen jährlich verarbeitet. (Stand 2013, nih.gov Announcement)

Um den Sucherfolg und die Spezifität der Ergebnisse zu verbessern, werden die Nutzereingaben durch einen Algorithmus verändert und erweitert. Der Dokumentation des „US National Institutes of Health“ zufolge, werden alle Eingaben, die nicht mittels korrekter Syntax markierte Suchfelder adressieren, mit dem „Automatic-Term-Mapping-Algorithmus“ bearbeitet. Dabei wird versucht, die Sucheingabe mit verschiedenen vorgegebenen Tabellen abzugleichen, um eine geeignete Datenbankabfrage zu generieren.

Zunächst wird in der Medical-Subject-Headings-Datenbank (MeSH) nach einer Übereinstimmung gesucht. Diese ist ein von der „US National Library of Medicine“ betriebener Thesaurus. Er setzt vereinheitlichte biomedizinische, chemische und publikationsspezifische Begriffe in Relation zueinander und ordnet sie in einer hierarchischen Struktur von Ober- und Unterüberschriften an. Bereits in MEDLINE®/PubMED® erfasste Artikel werden mit den passenden Kategorien verknüpft. Wird nach einer Eingabe, die als MeSH-Überschrift geführt ist, gesucht, so wird automatisch auch nach allen in der Hierarchie darunter stehenden Begrifflichkeiten gesucht.

Sollte in MeSH kein Treffer erfolgen, wird nun in der „Journals Translation Table“, einem Verzeichnis von wissenschaftlichen Journals, ihren Abkürzungen und Seriennummern gesucht.

Wird auch hier kein Treffer erzielt, so werden Tabellen mit den Namen von Autoren und Mitarbeitern durchsucht, die in MEDLINE® geführt werden.

Sollte auch dies nicht zum Erfolg führen, so wird die Eingabe in einzelne Wörter zerlegt, die dann in allen Feldern mit einer AND-Verknüpfung gesucht werden.

Mit Sternchen (*) markierte Eingaben werden als zu vervollständigender Wortstamm betrachtet. PubMed generiert für jede solche Zeichenkette bis zu 600 Variationen, die allerdings nicht vom „Automatic-Term-Mapping-Algorithmus“ weiter aufgearbeitet werden.

Daher kann es wichtig sein, zusätzlich nach Oberbegrifflichkeiten in einer gesonderten Eingabe zu suchen, obwohl bereits nach einem Wortstamm, der diese bereits erfasst haben sollte, gesucht wird.

2. Material und Methoden

2.1 Die systematische Literaturrecherche

Die Literaturrecherche und ihre Auswertung wurde vom Autor selbst durchgeführt. Dabei erhielt er unterstützende Hilfestellung von Prof. Dr. Sebastian Straube (damals Institut für Arbeits-, Sozial- und Umweltmedizin, Universitätsmedizin Göttingen).

Ziel der entwickelten Suche und der dazugehörigen Verknüpfungen war es, möglichst alle Ressourcen der Datenbank auszuschöpfen, um alle relevanten verzeichneten Publikationen, die arbeitsbezogene Endpunkte in randomisierten, klinischen Studien zu chronischen Schmerzerkrankungen berichten, zu erfassen. Da viele rheumatische Erkrankungen mit chronischen Schmerzen einhergehen, wurden auch Studien zu schmerzhaften rheumatischen Erkrankungen in das vorliegende Review einbezogen.

Die Benutzereingabe wurde in 3 Begriffsgruppen unterteilt (Abbildung 1). Die Begriffe innerhalb der Gruppen wurden mit einer ODER-Verknüpfung versehen, während die Gruppen mit UND-Verknüpfungen verbunden wurden.

Die erste Gruppe umschreibt die arbeitsbezogenen Endpunkte. Es wurde nach Wortstämmen und Synonymen für eine Erwerbstätigkeit gesucht: „work“, „work*“, „employment“, „employ*“, „job“, „job*“, „labor“, „labor*“, „labour*“ (British English). Weiter wurde nach allgemein ökonomisch orientierten Studien gesucht („economic“) sowie nach Begriffen, die sich mit verlorener Arbeitszeit oder Krankschreibung beschäftigen: „LWDE“ (*lost workday equivalents*), „sick leave“, „sick listed“.

Die Begriffe „workday“ und „lost workday equivalents“ aus vorherigen Suchen zu Beginn der Recherche für diese Dissertation (Wolf et al. 2012) wurden in Ermangelung eines MeSH-Eintrags im Wortstamm „work“ zusammengefasst.

Die zweite Gruppe beschäftigte sich mit Begriffen, die mit chronischen Schmerzen im Zusammenhang stehen. Neben allgemeinen Sucheingaben wie „joint“, „joint*“, „chronic AND pain“, „chronic AND pain*“, „back pain“, „neck pain“, „neuralgia“, „myalgia“ wurden auch spezifische Diagnosen abgefragt. Anhang 3 enthält eine vollständige Übersicht aller erfassten Erkrankungen. Die zuvor als relevant

erachteten Begriffe (Wolf et al. 2012) „chronic myofascial pain“ und „postherpetic neuralgia“ wurden wegen Redundanz mit „neuralgia“ und „myofascial pain syndrome“ entfernt.
In der letzten Gruppe wird die Studienart definiert. Nur arbeitsbezogene Endpunkte, die mit Placebo und Scheininterventionen (Sham) kontrolliert wurden, sollten erfasst werden. Die Beschränkung auf kontrollierte, klinische Studien sollte die Vergleichbarkeit der Ergebnisse, vor allem in Hinblick auf eine spätere Zusammenfassung in Meta-Analysen, gewährleisten.

Um alle möglichen Quellen an Informationen zu nutzen und um einen systematischen Fehler zu vermeiden, wurde die Suche ohne Anwendung von Filtern durchgeführt.
Die Suche erfolgte ein weiteres Mal ohne den arbeitsbezogenen Teil der Sucheingabe, um die Häufigkeit abzuschätzen, mit der arbeitsrelevante End-punkte in randomisierten, placebokontrollierten Studien berichtet wurden. So konnte auf eine Grundgesamtheit von Studien geschlossen werden, die sich mit chronischen Schmerzerkrankungen und rheumatischen Erkrankungen befassten und die arbeitsbezogene Endpunkte untersucht und berichtet haben könnten.

Damit der volle Vorteil aus dem „Automatic-Term-Mapping“ gezogen werden konnte, wurden die Begriffe „work“, „labor“, „employment“, „job“, „pain“, „joint“ als Einzeleingaben unabhängig von ihren Wortstämmen formuliert. Eine Aufstellung aller durch Mapping erzeugten Sucheingaben findet sich in Anhang 4.

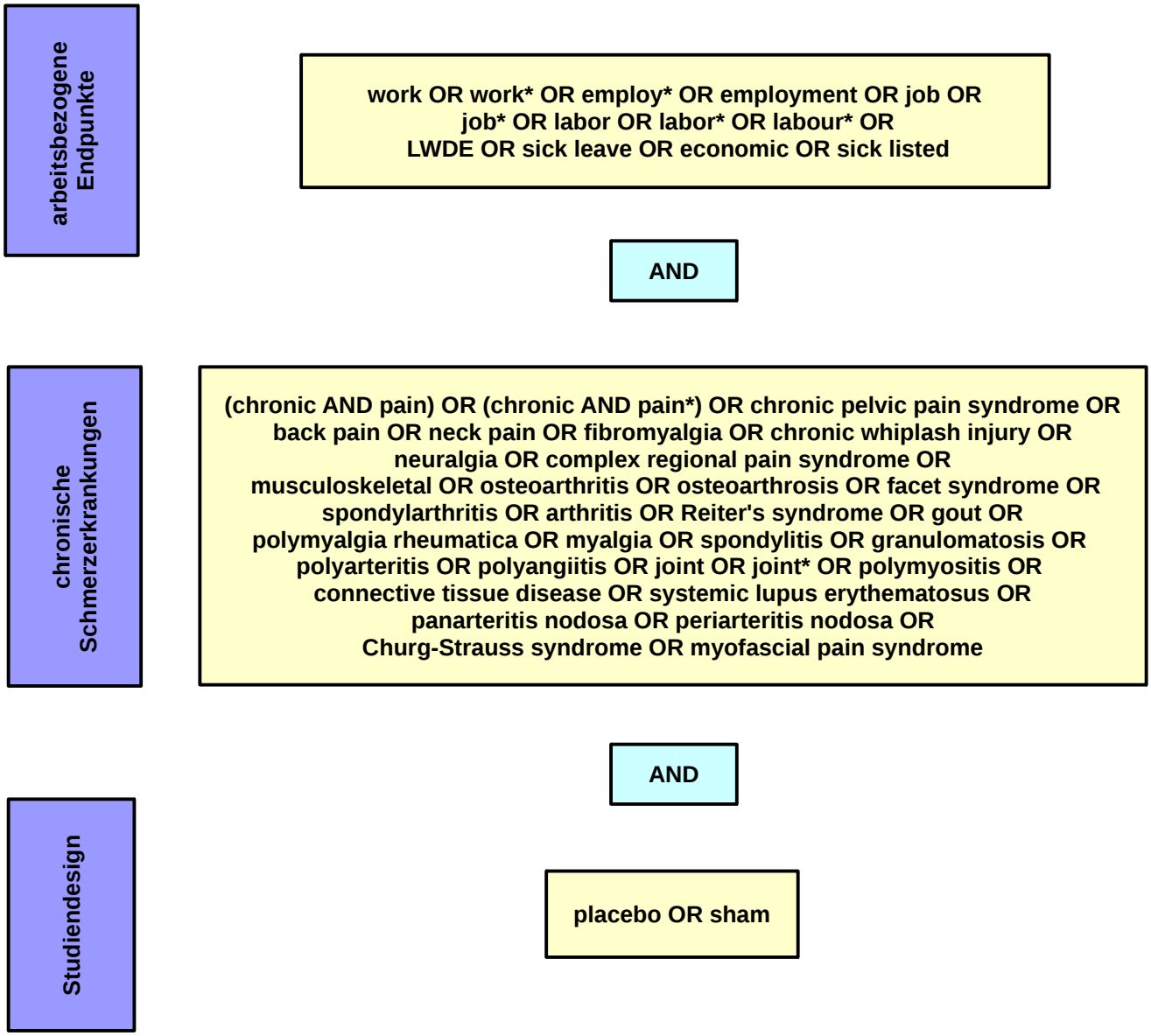


Abbildung 1: *Pubmed-Sucheingabe*
LWDE = lost workday equivalents

Die gefundenen Datenbankeinträge wurden zuerst im Titel und Abstract nach arbeitsbezogenen Themen durchsucht. Potenziell relevante Studien wurden dann im Volltext gesichtet. Wegen der sprachlichen Fähigkeiten des Autors wurden nur in englischer oder deutscher Sprache verfasste Artikel ausgewertet.

Es wurden nur solche Studien als geeignet befunden, die für sich in Anspruch nahmen klinisch, randomisiert und mindestens einfachblind mit Placebo

kontrolliert zu sein und in irgendeiner Form über arbeitsbezogene Endpunkte berichteten. Weiterhin sollten dabei Patienten erfasst worden sein, die entweder seit mindestens 3 Monaten an den meisten Tagen der Woche an Schmerzen litten, oder von einer rheumatischen Erkrankung betroffen waren.

2.2 Bewertung der Studienqualität

Um einen Eindruck von der Durchführung beziehungsweise der Beschreibung der gefundenen Studien zu bekommen, wurde die „Oxford Quality Scale“ (OQS), veröffentlicht von Jadad et al. 1996, verwendet.
Die OQS besteht aus drei Fragen, die sich auf die Berichterstattung der verwendeten Studienmethodik konzentrieren.
Die erste Frage dient der Beurteilung der Patientenrandomisierung. Für eine nicht randomisierte Studie, oder eine falsch vorgenommene Randomisierung werden 0 Punkte vergeben. Die einfache Erwähnung, dass eine Studie randomisiert sei ohne eine genauere Beschreibung ihrer Durchführung, wird mit einem Punkt bewertet. Zwei Punkte bekommt die Studie, wenn die Art der Randomisierung beschrieben und für angebracht erachtet wurde.
Die zweite Frage bezieht sich auf das Vorhandensein von Doppelverblindung. Sie verhält sich von der Bewertung her analog zur ersten Frage.
Die abschließende dritte Frage untersucht, ob über die Anzahl und die Gründe der Studienabbrecher berichtet wurde. Sie kann mit maximal einem Punkt honoriert werden.
So kann eine Studie als höchste Bewertung eine 5 und als niedrigsten Wert eine 0 erreichen.
Für diese Arbeit stand ein systematisches Untersuchen der Studien sowie deren Dokumentation im Vordergrund. Da ein möglichst weit gefasster Überblick über arbeitsbezogene Endpunkte in der vorliegenden Arbeit dargestellt werden sollte, wurden keine Studien basierend auf ihrer Bewertung von dem systematischen Review ausgeschlossen. Auch wurden die OQS-Wertungen nicht als

Gewichtungen für die vom Autor erstellten Meta-Analysen verwendet. Eventuelle Einflüsse der Studienqualität auf die Ergebnisse wurden mittels Regressionen untersucht, um zu ergründen, ob eine höhere Studienqualität mit geringeren Behandlungseffekten assoziiert ist. Dabei wurden die quadrierten Standardfehler der standardisierten Mittelwertsdifferenzen als Gewichte verwendet.

2.3 Datenextraktion

Die genauen Ergebnisse der Studien wurden auf vielfältige Weise berichtet. Die meisten Studien gaben ihre kontinuierlichen Endpunkte entweder als Differenz zwischen Studienbeginn und -ende oder als absolute Werte zu Beginn und Ende der Studie an. Letzteres erschwerte eine Meta-Analyse, da man mit den gegebenen Durchschnittswerten ohne weiteres deren Differenz bilden kann, es jedoch nicht möglich ist, die zugehörigen Standardfehler beziehungsweise die Standardabweichung (s) zu errechnen, die zur Abschätzung der Präzision der Studie und zur Berechnung des Konfidenzintervalles (KI) zwingend notwendig sind. Das arithmetische Mittel der Stichprobenwerte diente als Schätzer für den durchschnittlichen Wert der Gesamtpopulation. n ist die Anzahl der Stichprobenwerte. Die Variabilität der gemachten Beobachtungen, also die Abweichungen der individuellen Werte (x) von dem Mittelwert ($\bar{x}$), wird Varianz (s^2) genannt. Sie ist die quadrierte Standardabweichung (Altman 1991).

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

Formel 1

n : *Anzahl der Stichproben*
x : *Stichprobenwerte*
$\bar{x}$: *Mittelwert*
s : *Standardabweichung*

Wie durch Formel 1 beschrieben, ist die Verfügbarkeit der individuellen Probandendaten zwingend erforderlich, um die Standardabweichung zu berechnen. Wurde die Standardabweichung der Mittelwertsdifferenzen in den Studien nicht berichtet, so wurden die Werte zu Studienende, sofern für diese

Werte die Standardabweichungen gegeben waren, separat von den Differenzwerten in Meta-Analysen vom Autor untersucht. Bei ausreichender Studiengröße sollte die Randomisierung der Patienten für die Homogenisierung der untersuchten Stichprobe ausreichen. Davor wurden die relevanten Baseline-Werte mit einem t-Test[2] auf statistisch signifikante Unterschiede geprüft.
Waren dagegen Konfidenzintervalle (Barkham et al. 2010) oder die p-Werte[3] (Straube et al. 2011) innerhalb der Behandlungsgruppen gegeben, so wurde die Standardabweichung der Differenz innerhalb der Gruppen unter Zuhilfenahme von Probandenanzahl und der t-Verteilung[4] geschätzt (Higgins und Green 2011).
Gaben die Studien Median und die 25%- und 75%-Quantile an (Albert et al. 2013, Kavanaugh et al. 2006, van der Heijde et al. 2006), so wurde versucht, den Mittelwert und dessen Standardabweichung anhand dieser zu schätzen. Dazu wurde der Median für den Mittelwert eingesetzt und es wurde weiterhin angenommen, dass der Abstand zwischen den Interquartilen 1,35 Standardabweichungen entspricht. Diese Annahmen sind jedoch nur gültig, wenn die Endpunktwerte annähernd normalverteilt sind (Higgins und Green 2011). Sie werden in Abschnitt 4.1 näher diskutiert.

Oftmals lagen die benötigten Daten nicht in Textform vor, sondern wurden als Grafik dargestellt (Egsmose et al. 1997, Kavanaugh et al. 2006, Kavanaugh et al. 2013, van der Heijde et al. 2006). Diese wurden von dem Autor dieser Dissertation vergrößert ausgedruckt und die Daten mithilfe von geometrischen Konstruktionen und Messungen abgeleitet.

---

2 t-Test: Statistischer Hypothesentest, der bei angenommener Normalverteilung der Daten zur Anwendung kommt

3 p-Wert: Ergebnis eines vorangegangenen Hypothesentests, welches die Wahrscheinlichkeit angibt, dass die getesteten oder extremeren Stichprobenwerte unter Annahme der Null-Hypothese auftreten

4 Studentsche t-Verteilung: Wahrscheinlichkeitsverteilung, welche bei normal verteilten Stichproben Anwendung findet

2.4 Statistische Verfahren

Alle statistischen Berechnungen wurden von dem Verfasser der vorliegenden Dissertation selbst durchgeführt. Beraten und unterstützt wurde er dabei von Prof. Dr. Tim Friede, Leiter des Instituts für Medizinische Statistik der Universitätsmedizin Göttingen.
Allen in dieser Arbeit erstellten Analysen wurde ein Signifikanzniveau[5] von 5% zu Grunde gelegt.

2.4.1 Meta-Analysen und Heterogenität

Um die Ergebnisse mehrerer Studien zu einem gepoolten Durchschnittswert zusammenfassen zu können, wurden in der vorliegenden Arbeit eine Reihe von Meta-Analysen erstellt. Sie wurden mit dem Programm Review Manager (RevMan) der Cochrane Collaboration berechnet (Version 5.2. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2012).
Um die Analyse der Daten zu vereinfachen, prüft das Programm automatisch auf Heterogenität der Ergebnisse mit einem Chi-Quadrat-Test. Higgins et al. erklären 2003, dass dabei die quadrierten Abweichungen vom kombinierten Mittelwert der Meta-Analyse mit der bereits in der Analyse verwendeten Gewichtung versehen und dann aufaddiert werden, um diese dann unter Berücksichtigung der Freiheitsgrade, die von der Anzahl der untersuchten Studien abhängen, mit einer Chi-Quadrat-Verteilung zu vergleichen. Der so erhaltene p-Wert kann als Indikation für Heterogenität der Studienergebnisse verwendet werden. Der Chi-Quadrat-Test sei jedoch zu sensitiv, wenn sehr viele Studien und Probanden in einer Analyse kombiniert würden, da eine gewisse Heterogenität zwischen den Studien unvermeidbar sei (Higgins et al. 2003). Gleichzeitig konnten Hardy und Thompson 1998 zeigen, dass bei geringer Probandenanzahl, oder wenn eine

5 Signifikanzniveau: Schwellenwert ab dem der Unterschied zweier Messgrößen als statistisch signifikant definiert ist. Es entspricht der Wahrscheinlichkeit für einen Fehler 1. Art, nachdem man die Nullhypothese, d.h. die Annahme, dass kein nur auf Zufall basierter Unterschied zwischen den Werten besteht, ablehnt, obwohl sie wahr ist.

einzelne Studie durch höhere Präzision und daraus folgender höherer Gewichtung die Meta-Analyse dominiert, die Teststärke (p*ower*), also die Wahrscheinlichkeit einen Unterschied zu finden, gering ist. Durch diesen Effekt würde bei vielen Meta-Analysen die Untersuchung der Heterogenität ergebnislos verlaufen.

Daher entwickelten Higgins et al. 2003 den Wert I^2 (Formel 2). Er leitet sich aus der Differenz der klassischen Chi-Quadrat-Statistik und ihren Freiheitsgraden ($k-1$) ab, die durch die Chi-Quadrat-Statistik Q dividiert wird und mit 100% multipliziert wird. k gibt dabei die Anzahl der untersuchten Studien an.

$$I^2 = 100\,\% \times \frac{Q-(k-1)}{Q}$$

Formel 2

$k: Anzahl\, der\, Studien$
$Q: Chi-Quadrat-Statistik$
$I^2: Heterogenität\ \text{in}\ Prozent$

Negative Werte sind als 0 % definiert. Positive Werte geben eine Schätzung des Anteils der Variation an, die auf Heterogenität der Studien zurückzuführen ist. Als grobe Orientierung ordneten Higgins et al. 2003 den Werten 25%, 50% und 75% die Beschreibungen niedrige, mittlere oder hohe Heterogenität zu.

Wird Heterogenität festgestellt, so ist zu entscheiden, wie man mit dieser umgeht und welches Meta-Analysen-Modell zum Einsatz kommt.

2.4.2 Modelle der Meta-Analysen

Den verwendeten Meta-Analysen liegen vornehmlich zwei Modelle zu Grunde: das *fixed effect model* (FE) und das *random effects model* (RE).

Das *fixed effect model* basiert auf der Annahme, dass es in der zu untersuchenden Population einen allgemeingültigen, homogenen Behandlungseffekt gibt. Die unterschiedlichen Ergebnisse der einzelnen Studien entstehen hierbei durch Ungenauigkeiten bedingt durch begrenzte Stichprobenziehung. Dementsprechend werden die Gewichtungen der einzelnen Studien gemäß der reziproken Varianz ihres geschätzten Behandlungseffektes vergeben. Diese

richtet sich, da aus dem Quadrat des Standardfehlers errechnet, primär nach der Studiengröße, kann aber auch unterschiedliche Probandenanzahl innerhalb der verschiedenen Behandlungsarme berücksichtigen (Deeks und Higgins 2010). Das Ergebnis des *fixed effect model* beschreibt einen allgemeinen Behandlungseffekt, dessen Schätzsicherheit mit einem 95%-Konfidenzintervall beschrieben wird (Higgins und Green 2011).

Bei dem *random effects model* wird angenommen, dass die verschiedenen zu Grunde liegenden gemessenen Behandlungseffekte wegen unbekannter systematischer Unterschiede nicht in allen Studien gleich sind, jedoch einer Verteilung folgen. Es wird weiterhin angenommen, dass diese verschiedenen Behandlungseffekte zufälliger Natur und durch die Normalverteilung beschreibbar sind. Die unterschiedlichen zu Grunde liegenden Behandlungseffekte streuen mit der Varianz Tau^2 um einen Durchschnittswert. Liegt eine hohe Interstudienvarianz vor, erhalten die kleinen Studien mehr Gewicht, als sie es bei einem *fixed effect model* erhalten hätten, da viele kleine Studien mehr über die Verteilung der Behandlungseffekte verraten, als eine große Studie. So gibt das gepoolte Ergebnis des *random effects model* eine Schätzung über den Durchschnittswert verschiedener Behandlungseffekte in verschiedenen Studien ab. Die Schätzsicherheit dieses Durchschnittswertes wird mit dem zugehörigen 95%-Konfidenzintervall angegeben. Liegt keine Heterogenität vor, ist die Varianz der verschiedenen zu Grunde liegenden Behandlungseffekte des *random effects model* Null. In diesem Falle gibt es nur einen allgemeinen Behandlungseffekt und beide Modelle kommen folglich zum gleichen Ergebnis (Higgins und Green 2011). In der vorliegenden Dissertation wurde das *random effects model* verwendet, wenn der Heterogenitätswert I^2 größer als Null war, wenn nicht, kam das *fixed effect model* zum Einsatz.

2.4.3 Funnel-Plots

Der Funnel-Plot bietet eine weitere Möglichkeit Heterogenität zu untersuchen. Zumeist findet dieses Verfahren Anwendung, wenn vermutet wird, dass

vornehmlich Studien mit signifikanten Ergebnissen veröffentlicht wurden. Der daraus resultierende Effekt auf Meta-Analysen wird Publikationsverzerrung (*publication bias*) genannt. In dieser Arbeit wurden Funnel-Plots verwendet, welche auf der X-Achse den Behandlungseffekt abbilden, während die Y-Achse den Standardfehler der einzelnen Studien darstellt. Die vertikale Linie markiert den geschätzten Behandlungseffekt, die diagonalen Linien repräsentieren sein 95%-Konfidenzintervall (Sterne und Egger 2001).

Die Ergebnisse der einzelnen Studien werden nun unter Berücksichtigung des gepoolten Effektes analysiert. Ein unsymmetrisches Erscheinungsbild lässt auf Verzerrungen oder Heterogenität schließen. Beispielsweise spräche hier eine stark linkslastige Verteilung der Studien, besonders wenn es sich um kleine Studien mit großem Standardfehler handelt, für einen *publication bias* (Abbildung 2).

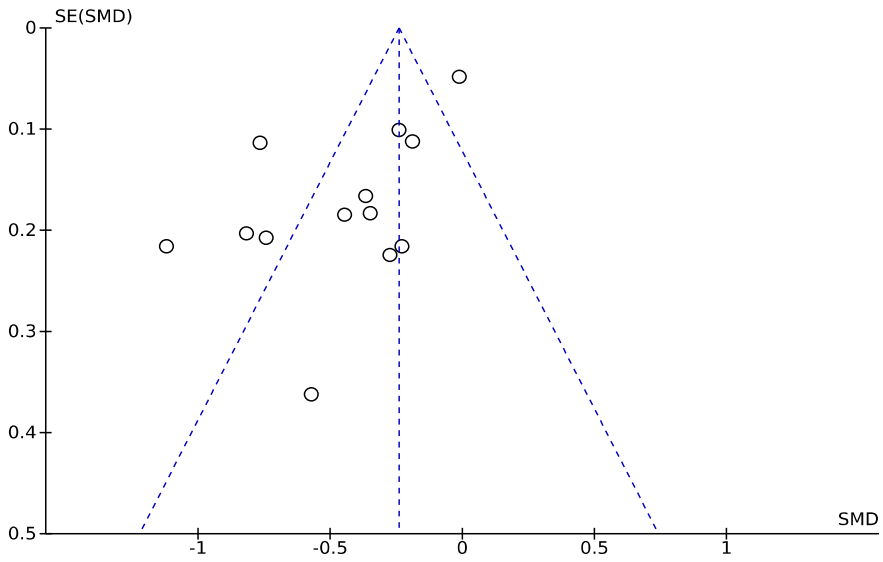


Abbildung 2: *Beispiel eines Funnel-Plot mit möglicher Publikationsverzerrung.*

SMD: standardisierte Mittelwertsdifferenz
SE(SMD): Standardfehler der standardisierten Mittelwertsdifferenz

Eine symmetrische Verteilung spricht gegen einen *publication bias* (Abbildung 3). Hopewell et al. 2009 konnten zeigen, dass Studien mit signifikanten Ergebnissen häufiger publiziert wurden als Studien mit unschlüssigen Ergebnissen. Publikationsverzerrungen stellen somit für Meta-Analysen ein reales Problem dar.

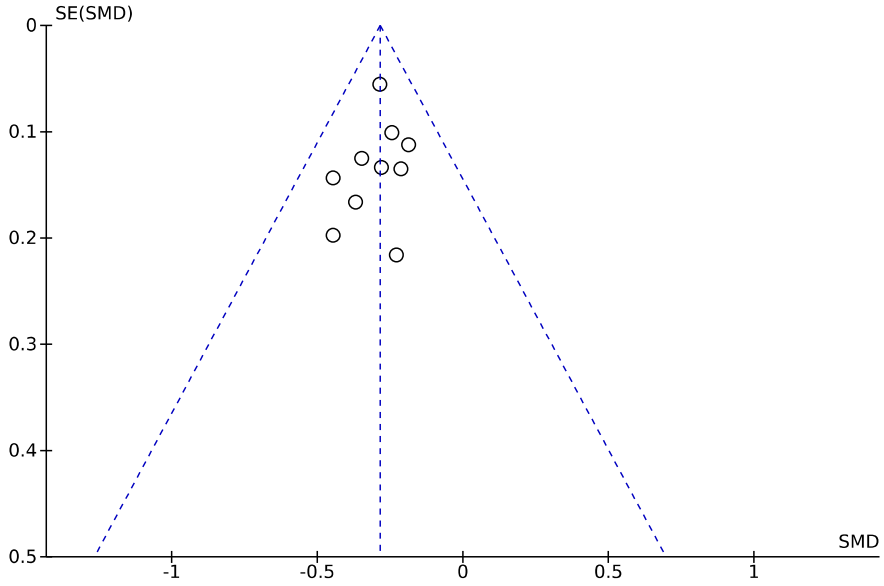


Abbildung 3: *Beispiel eines Funnel-Plot ohne Publikationsverzerrung*

2.4.4 Standardisierung der Studienergebnisse

Die ausgewerteten Studien verwendeten viele verschiedene Messinstrumente, um arbeitsbezogene Studienendpunkte zu untersuchen. Damit trotzdem die Ergebnisse zusammengefasst werden konnten, mussten die einzelnen Durchschnittsdifferenzen zwischen den untersuchten Behandlungsgruppen standardisiert werden. In dem Programm RevMan 5.2 kommt die Standardisierung nach Hedges zum Einsatz, auch bekannt als Hedges' g (Formel 3) (Higgins und Green 2011). Dabei werden die bei den beiden

Behandlungsgruppen errechneten Durchschnittswerte (m_1, m_2) subtrahiert und durch ihre gemeinsame Standardabweichung (s) dividiert. Um Verzerrungen bei kleineren Gruppengrößen zu vermeiden, wird mit einem Korrekturfaktor ($1-\frac{3}{4N-9}$) multipliziert, bei dem N die Summe der Patienten in beiden Behandlungsgruppen ist. (Hedges 1981).

$$SMD=\frac{m_1-m_2}{s}\times\left(1-\frac{3}{4N-9}\right)$$

Formel 3

N : *Anzahl der Patienten beider Gruppen*
m_1 : *Mittelwert Gruppe* 1
m_2 : *Mittelwert Gruppe* 2
s : *gepoolte Standardabweichung aus Gruppe* 1 *und* 2

Das erhaltene Ergebnis wird als standardisierte Mittelwertsdifferenz bezeichnet (engl. standardised mean difference, SMD). Sie drückt den Behandlungseffekt als Vielfaches der Standardabweichung aus.
Um diese unhandliche Einheit besser einordnen zu können, wurden an einigen Stellen der Arbeit die SMD-Ergebnisse mit den Standardabweichungen des betreffenden Endpunktes multipliziert, so dass ein Ausdrücken des Ergebnisses in den Einheiten des Endpunktes möglich ist. Die Einschränkungen dieser Vorgehensweise werden in Abschnitt 4.2.1.2 näher diskutiert.

2.4.5 Präsentation der Meta-Analysen

Präsentiert werden die im Rahmen dieser Arbeit erstellten Meta-Analysen mithilfe eines Forest-Plots, der es dem Leser ermöglicht Durchschnittswerte, Gewichtungen und Konfidenzintervalle der einzelnen Studien sowie der gesamten Meta-Analyse zu begutachten (Lewis und Clarke 2001).
In den gezeigten Grafiken stehen hierbei die Bezeichnungen „Mean Difference“ für „Mittelwertsdifferenz“ und „Std. Mean Difference“ für standardisierte Mittelwertsdifferenz“. Das Kürzel „IV“ steht für „inverse Varianz“, die hierbei verwendete Art der Gewichtung. Die Begriffe „Fixed“ und „Random“ zeigen an, ob

das *fixed effect model* oder das *random effects model* für die Meta-Analyse verwendet wurde. In allen Meta-Analysen wurden 95%-Konfidenzintervalle berechnet (95% CI/95%-KI).

2.4.6 Qualitative Endpunkte

Für die Analyse der Arbeitsfähigkeit in dem Abschnitt 3.4.7 wurden die von den Studien definierten Kategorien in ein binäres System überführt (arbeitsfähig ja oder nein). So konnten der Informationsgehalt und die Fragestellung vereinfacht werden. Dies kam der Homogenität und den verwendeten statistischen Verfahren entgegen. Oft benutzten die Studien verschiedene Kategorien, die nicht ohne weiteres miteinander zu vergleichen waren. Daher wurden alle Patienten, für die in dieser Arbeit durchgeführten Analysen, als arbeitsfähig eingestuft, wenn die Studie über Arbeitsfähigkeit, Beschäftigung in Vollzeit oder Halbzeit berichtete. Entsprechend erfolgte eine Einteilung als arbeitsunfähig bei Schlüsselwörtern wie Krankschreibung und ähnlichen Begriffen. Letztendlich wurde ein gepoolter Wahrscheinlichkeitsquotient (*risk ratio*, RR) berechnet. Dieser gibt das Verhältnis zwischen zwei Wahrscheinlichkeiten für ein bestimmtes Ereignis an. In dieser Arbeit wird sowohl betrachtet, wie viele arbeitsfähige Personen im Verlauf der Studien arbeitsunfähig wurden, als auch in einer zweiten Analyse, wie viele arbeitsunfähige Personen wieder arbeitsfähig wurden. Die Wahrscheinlichkeit für das jeweilige Ereignis der aktiven Gruppe wurde dann durch die Wahrscheinlichkeit der Placebogruppe dividiert. Der Vorteil dieser Darstellungsweise liegt darin, dass ein direkter Vergleich zwischen aktiver Behandlung und Placebo-Behandlung möglich ist. Der so berichtete Wert sagt jedoch nichts über die absoluten Wahrscheinlichkeiten eines Ereignisses aus. Um die Größe des Behandlungseffektes für das einzelne Individuum besser einschätzen zu können, ist es hilfreich, ebenfalls das durchschnittliche Risiko der Placebogruppe sowie die Wahrscheinlichkeitsdifferenz (*risk difference*, RD) der beiden Gruppen zu kennen (Higgins und Green 2011). Bei der

Wahrscheinlichkeitsdifferenz werden die Anteile der eingetroffenen Ereignisse der Kontrollgruppe von dem Anteil der aktiven Gruppe subtrahiert.
Die Ansprechraten der schmerzbezogenen und rheumatischen Endpunkte in Abschnitt 3.6 wurden ebenfalls mithilfe von RR und RD untersucht. Dazu wurde die absolute Ereignisanzahl von den angegebenen Ansprechraten und den jeweiligen Gruppengrößen abgeleitet.

2.4.7 Trends und Regressionen

Die Regressionen betreffenden statistischen Berechnungen erfolgten mit dem Programm R: A language and environment for statistical computing Version 3.0.3 (The R Foundation for Statistical Computing, R Core Team (2014), R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/). Die Organisation, die Aufbereitung und die für das Einlesen der Daten notwendige Formatierung in Textdateien wurde mit Libre Office 4.1 durchgeführt.
Bei der Untersuchung der jährlichen Veröffentlichungszahlen der relevanten Artikel wurde die Entwicklung der Grundgesamtheit der Studien, die sich mit chronischen Schmerzen beschäftigten, durch Logarithmierung auf einen möglichen exponentiellen Trend hin untersucht. Der tatsächliche Fit (Regressionsgerade), welcher den Anstieg der Veröffentlichungen beschreibt, wurde dann mithilfe einer exponentiellen Funktion mit drei Parametern erstellt (Formel 4).

$$Datenbanktreffer = e^{\frac{x - 1951}{b}} \times \frac{1}{d} - c$$

x : Jahre
b,d : Multiplikationsfaktoren
c : Ordinatenabschnitt

Formel 4

Hier beschreibt x die Veröffentlichungsjahre, während der Ordinatenabschnitt (c) und die Multiplikationsfaktoren (b, d) dazu dienen, den Kurvenverlauf flacher zu gestalten. Um die Analyse zu vereinfachen, wurden die Jahre durch die Subtraktion von 1951 auf 1 und ansteigend reduziert, da die erste erfasste Arbeit im Jahr 1952 veröffentlicht wurde.

Alle weiteren Trends wurden mit linearen Modellen beschrieben, die sich aus zwei Parametern, nämlich der Steigung und dem Ordinatenabschnitt zusammensetzen. Bei Regressionen wird mit der Methode der kleinsten Quadrate das Modell mit den geringsten horizontalen, quadrierten Abweichungen (Residuen) gesucht. Um die Aussagekraft der Regression einschätzen zu können, wird die Varianz der abhängigen Variable durch die Varianz der Residuen dividiert. Der Quotient kann nun mit der F-Verteilung[6] verglichen werden, um einen p-Wert zu erhalten (Altman 1991).
Bei den Regressionen, die einzelne Studienergebnisse verarbeiteten (Abschnitte 3.4.1.2; 3.5.2; 3.7), kam die inverse Varianz der abhängigen Variablen, die den quadrierten Standardfehlern der standardisierten Mittelwertsdifferenzen entsprechen, als Gewichtung zum Einsatz.

2.4.8 „Votecount“

In Abschnitt 3.4.1 findet sich eine Übersicht, geteilt nach erreichter statistischer Signifikanz der Studienergebnisse. Dieser „Votecount“ kann dem Leser eine kurze Orientierung über die gemachten Beobachtungen der erfassten Studien bieten (Rodgers et al. 2009). Für die Berechnung der Wahrscheinlichkeit, ob der durchschnittliche Behandlungseffekt eine statistisch signifikante Verbesserung der untersuchten Endpunkte ergab, wurde der Anteil der als Erfolg gewerteten Studien als Schätzer herangezogen. Das Konfidenzintervall wurde mit der Methode von Clopper und Pearson 1934 unter Verwendung des Statistikprogrammes R 3.0.3 bestimmt.

6 F-Verteilung: Wahrscheinlichkeitsverteilung, die häufig für Hypothesentests in Bezug auf Varianzanalysen verwendet wird

3. Ergebnisse

3.1 Studienübersicht

Die Datenbankabfrage ergab 1063 potenziell relevante Veröffentlichungen aus den Jahren von 1958 bis 2013 (Datum der letzten Suche 13.06.2013).

Von den 1063 Treffern wurden 948 nach Durchsicht des Abstracts als nicht relevant ausgeschlossen.
Nach Sichtung des Volltextes der übrigen 115 Veröffentlichungen konnten 30 identifiziert werden, die sich mit placebokontrollierten Studien befassten sowie über arbeitsbezogene Endpunkte berichteten und somit die gesetzten Kriterien (Abschnitt 2.1) erfüllten. Die 85 nach Sichtung des Volltextes ausgeschlossenen Studien und die Gründe für ihren Ausschluss sind in Anhang 2 gelistet. In diesem Zusammenhang konnte nach Durchsicht des Reviews von Khadilkar et al. 2008 eine weitere relevante Studie gefunden werden (Jarzem et al. 2005), so dass insgesamt 31 für das systematische Review geeignete Arbeiten identifiziert wurden (Abbildung 4).

Die Veröffentlichungen von Arnold et al. 2005 und Chappell et al. 2008 wurden von Arnold et al. 2009 mit zwei weiteren Studien in einer Meta-Analyse kombiniert. Die Primärpublikationen berichteten jedoch über andere Endpunkte als das Review und fanden daher als Einzelstudien ebenfalls Berücksichtigung. Des weiteren veröffentlichte die betreffende Arbeitsgruppe, in der Arbeit Bradley et al. 2010, ein weiteres Review über dieselben Studien.
Die Veröffentlichungen von Bennett et al. 2003 und 2005 werden, da sie dieselbe Studie beschreiben, der Übersicht halber nur in Kombination erwähnt.
Straube et al. 2011 erstellten eine Meta-Analyse mithilfe von Rohdaten, welche in den Studien von Arnold et al. 2008, Crofford et al. 2005, Mease et al. 2008 und Pauer et al. 2008 erhoben wurden.
Somit ergeben sich 27 voneinander unabhängige Patientenkollektive.

In diesen wurden insgesamt 11434 Patienten untersucht. Die Studiendauer variierte zwischen 3 Wochen und 24 Monaten.

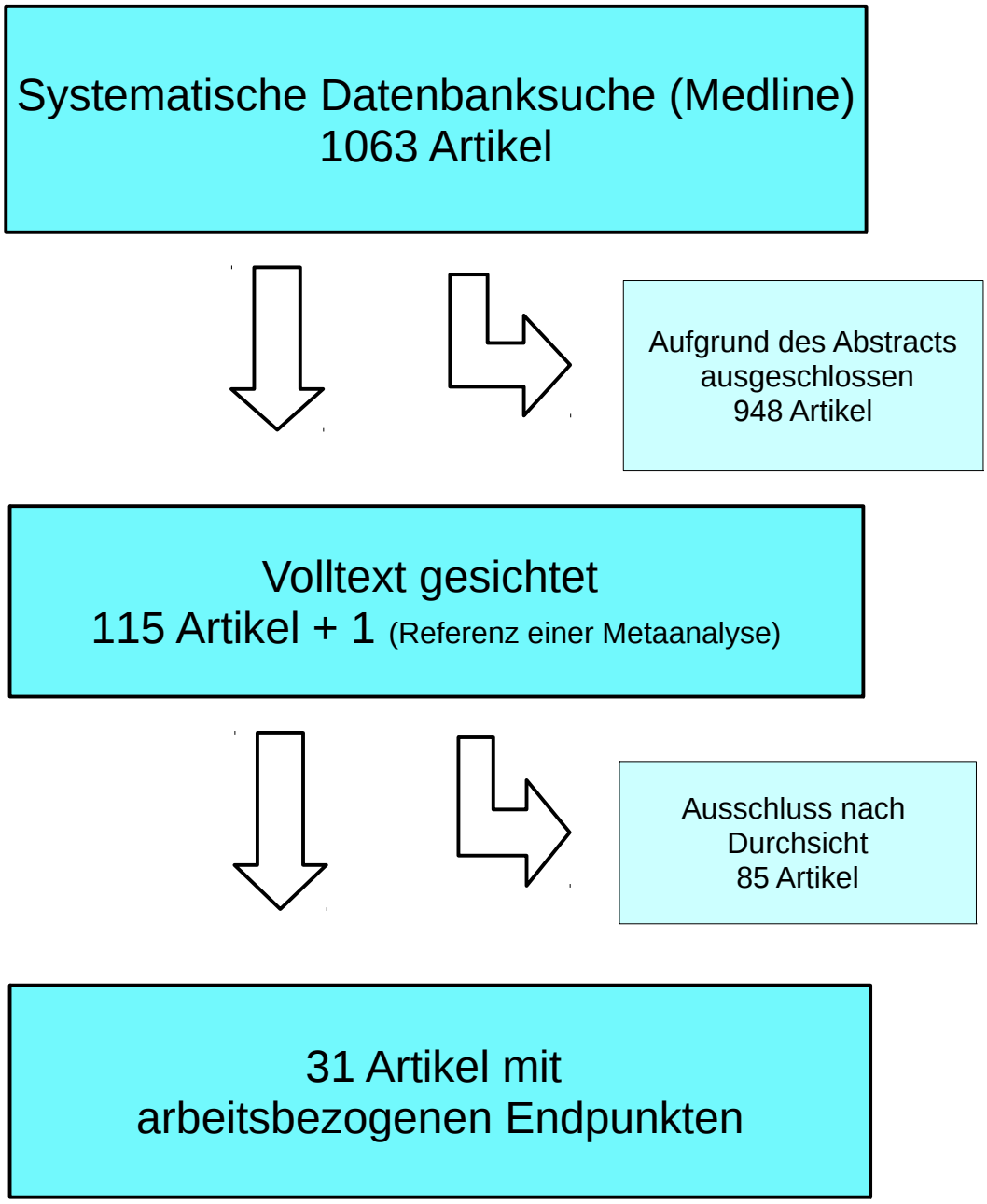


Abbildung 4: *Flowchart der systematischen Literaturrecherche*

Von den 31 in das systematische Review eingeschlossenen Arbeiten untersuchten 21 Studien Patienten mit chronischen Schmerzen.

- Neun beschäftigten sich mit chronischen Rückenschmerzen
  (Albert et al. 2013, Baron et al. 2010, Carlsson und Sjölund 2001, Jarzem et al. 2005, Lehmann et al. 1986, Licciardone et al. 2003, Skljarevski et al. 2009, Skljarevski et al. 2010a, Skljarevski et al. 2010b)
- Sieben mit Fibromyalgie
  (Arnold et al. 2005, Arnold et al. 2009, Bennett et al. 2003 & 2005, Bradley et al. 2010, Chappell et al. 2008, Straube et al. 2011)
- Zwei mit Osteoarthritis (Chappell et al. 2009, Chappell et al. 2011)
- Zwei mit chronischen Nackenschmerzen
  (He et al. 2005, Manchikanti et al. 2010)
- Van der Heijde et al. 2006 untersuchten Spondylitis ankylosans

Dabei wurde neun Mal über die Behandlung mit Duloxetin, vier Mal über Elektrostimulation oder Akupunktur, zwei Mal über Pregabalin und je ein Mal über ein Paracetamol-Tramadol-Kombipräparat, Amoxicillin-Clavulansäure, manuelle Osteopathieintervention oder Bupivacain-Injektionen sowie einmal über Infliximab berichtet. Zwei der Studien behandelten ihre Patienten zusätzlich zu den aktiven oder Placebotherapien mit weiteren Interventionen. Im weiteren Verlauf der Arbeit werden diese als Add-on-Therapie-Studien bezeichnet. So injizierten Manchikanti et al. 2010 Bupivacain entweder mit oder ohne Steroidzusatz, während Lehmann et al. 1986 zusätzlich zu den aktiven und funktionslosen Geräten der Reizstromtherapien alle Patienten mit Rückengymnastik behandelten.

Die verbliebenen zehn Studien untersuchten Patienten, die nicht explizit seit drei Monaten über chronische Schmerzen klagten, aber an einer schmerzhaften, rheumatisch Erkrankungen litten:

- Fünf Studien über rheumatoide Arthritis
  (Bejarano et al. 2008, Kavanaugh et al. 2009a, Meireles et al. 2010, Smolen et al. 2006, Strand et al. 1999)

- Zwei über Psoriasisarthritis
  (Kavanaugh et al. 2006, Kavanaugh et al. 2013)
- Eine über reaktive Arthritis (Egsmose et al. 1997)
- Eine über Spondylitis ankylosans (Barkham et al. 2010)
- Eine über Osteoarthritis (Markenson et al. 2005)

Dabei wurde zwei Mal mit Infliximab sowie je ein Mal mit Oxicodon, Leflunomid oder Methotrexat, Entanercept, Sulfasalazin, Certolizumab, Adalimumab mit Methotrexat, Golimumab und mit Niedrigenergielaser therapiert.
Drei der zehn Studien verwendeten ein Addon-Konzept. Kavanaugh et al. 2009a behandelten alle Probanden mit Methotrexat und zusätzlich entweder mit Certolizumab oder einem entsprechenden Placebo. Ebenso verfuhren die Gruppen um Smolen et al. 2006 mit ihrer Infliximab-Studie und Bejarano et al. 2008 mit ihrer Adalimumab-Studie. Tabelle 1 listet eine Übersicht der untersuchten Studien.

Patientenkollektive	Studie	Erkrankung	Aktive Behandlung	Probandenstärke	arbeitsbezogene Endpunkte	Studiendauer	OQS
1	Albert et al. 2013	chr. Rückenschmerzen	Amoxicillin-Clavulansäure	162	Arbeitszeitverluste	52 w	5
2	Baron et al. 2010	chr. Rückenschmerzen	Pregabalin	217	WPAI	5 w	4
3	Carlsson & Sjölund 2001	chr. Rückenschmerzen	Akupunktur, Elektroakupunktur	51	Beschäftigungsstatus	32 w	5
4	Jarzem et al. 2005	chr. Rückenschmerzen	TENS	350	MacGill Work Scale	12 w	5
5	Lehmann et al. 1986	chr. Rückenschmerzen	TENS, Elektroakupunktur	54	Beschäftigungsstatus	3 w	4
6	Licciardone et al. 2003	chr. Rückenschmerzen	manuelle Osteopathiebehandlung	91	Arbeitszeitverluste	20 w	3
7	Skljarevski et al. 2009	chr. Rückenschmerzen	Duloxetin	404	BPI-I	13 w	5
8	Skljarevski et al. 2010a	chr. Rückenschmerzen	Duloxetin	236	BPI-I, WPAI, Arbeitszeitverluste	13 w	4
9	Skljarevski et al. 2010b	chr. Rückenschmerzen	Duloxetin	401	BPI-I, WPAI, Arbeitszeitverluste	12 w	3
10	He et al. 2005	chr. Nackenschmerzen	Akupunktur & Elektrostimulation & Ohr-Akupressur	24	Einschränkungen am Arbeitsplatz	4 w	2
11	Manchikanti et al. 2010	chr. Nackenschmerzen	Bupivacain mit Steroid Injektionen	120	Arbeitsfähigkeit, Beschäftigungsstatus	96 w	5
12	Bennett et al. 2003 & 2005	Fibromyalgie	Paracetamol-Tramadol	315	FIQ, Arbeitszeitverluste, SF-36	8 w	3
13	Bradley et al. 2010 & Arnold et al. 2009 (Arnold et al. 2005 & Chappell et al. 2008)	Fibromyalgie	Duloxetin	1332	Bradley 2010: FIQ Arnold 2009: SDS (Arnold 2005: BPI) (Chappell 2008: SDS)	12-28 w	Review (Arnold 2005: 5) (Chappell 2008: 3)
14	Straube et al. 2011	Fibromyalgie	Pregabalin	2757	FIQ, Arbeitszeitverluste SF-36, SDS, MAF	8-14 w	Review
15	Chappell et al. 2009	Osteoarthritis/Knie	Duloxetin	231	BPI-I	13 w	5
16	Chappell et al. 2011	Osteoarthritis/Knie	Duloxetin	256	BPI-I	13 w	5
17	Markenson et al. 2005	Osteoarthritis	Oxicodon	109	BPI-I	12 w	4
18	Kavanaugh et al. 2006	Psoriasisarthritis	Infliximab	200	SF-36, Beschäftigungsstatus, Arbeitsfähigkeit, VAS-Produktivität, Arbeitszeitverluste	22 w	3
19	Kavanaugh et al. 2013	Psoriasisarthritis	Golimumab	405	VAS-Produktivität	24 w	4
20	Egsmose et al. 1997	reaktive Arthritis	Sulfasalazin	83	Arbeitszeitverluste	24 w	3
21	Bejarano et al. 2008	rheumatoide Arthritis	Adalimumab & Methotrexat	148	Beschäftigungsstatus, WIS, Arbeitszeitverluste	56 w	5
22	Kavanaugh et al. 2009a	rheumatoide Arthritis	Certolizumab & Methotrexat	1601	WPS-RA, Arbeitszeitverluste	24 & 52 w	3
23	Meireles et al. 2010	rheumatoide Arthritis	Niedrigenergielaser	82	DASH	8 w	5
24	Smolen et al. 2006	rheumatoide Arthritis	Infliximab + Methotrexat	1004	Arbeitsfähigkeit, Arbeitszeitverluste	54 w	5
25	Strand et al. 1999	rheumatoide Arthritis	Leflunomid, Methotrexat	482	Arbeitsproduktivität	48 w	4
26	Barkham et al. 2010	Spondylitis ankylosans	Entanercept	40	WIS, Arbeitszeitverluste	12 w	3
27	van der Heijde et al. 2006	Spondylitis ankylosans	Infliximab	279	SF-36, VAS-Produktivität, Arbeitszeitverluste	24 w	4

BPI-I = Brief Pain Inventory: Interference with normal work; chr. = chronisch; DASH = Disabilities of the Arm, Shoulder and Hand Questionnaire; FIQ = Fibromyalgia Impact Questionnaire; MAF = Multidimensional Assessment of Fatigue; OQS = Oxford Quality Scale; SDS = Sheehan Disability Scale; SF-36 = Short Form 36 Survey; TENS = Transcutaneous Electrical Nerve Stimulation; VAS = Visual Analog Scale; WIS = Work Instability Scale; WPAI= Work Productivity and Activity Impairment Questionnaire; WPS-RA = Rheumatoid Arthritis Sspecific Work Productivity Survey

Tabelle 1: *Studieneigenschaften*

Im Rahmen dieses Reviews wurden von den 31 Veröffentlichungen 27 Primärpublikationen mit der Oxford Quality Scale bewertet. Da die OQS nur für Primärpublikationen geeignet ist, wurden die 3 Reviews von Arnold et al. 2009, Bradley et al. 2010 und Straube et al. 2011 nicht bewertet. Die Arbeiten von Bennett et al. 2003 & 2005, die gleiche Untersuchung betreffend, werden hier als eine Publikation betrachtet. Von den bewerteten Veröffentlichungen wies der größte Teil eine hohe methodologische Studienqualität auf. Elf erhielten den maximal möglichen Wert 5/5. Sieben Arbeiten wurden mit 4/5 bewertet. Acht Arbeiten wurden mit 3/5 bewertet. Die Studie von He et al. 2005 erhielt nur 2/5 Punkten.

In jeweils zwölf Arbeiten wurden Punkte in den Kategorien Randomisierung und Verblindung verloren. Zumeist führte hier mangelnde Berichterstattung der verwendeten Verfahren zu niedrigen Bewertungen. Bei den Studien von He et al. 2005 und Licciardone et al. 2003 war keine Doppelverblindung anzunehmen. Daher verloren sie jeweils zwei Punkte in der betreffenden Kategorie. Die Studie von Kavanaugh et al. 2006 berichtete als einzige nicht über eventuelle Studienabbrecher und verlor hier dementsprechend einen Punkt.

Anhang 1 enthält eine genaue Beschreibung der einzelnen Studieneigenschaften.

Ergänzend sind die Charakteristika der in den Patientenkollektiven untersuchten Patienten in Tabelle 2 zusammengefasst.

	Patientenkollektive	Median	25-75% Interquantile	Spannweite (Min-Max)
Patientenanzahl	27	231	100,0-402,5	24-2757
weiblich (%)	27	65,1	51,1-75,4	19,3-100
Durchschnittsalter (Jahren)	27	49,8	46,0-52,7	34,2-63
Mittlere Dauer der chronischen Schmerzerkrankung (Jahre)	19	8,6	6,9-9,4	0,9-15,5

Tabelle 2: *Probandencharakteristika*

3.2 Untersuchung von arbeitsbezogenen Endpunkten in der Literatur

Die Grundgesamtheit aller Veröffentlichungen, die sich mit chronischen Schmerzerkrankungen oder rheumatischen Erkrankungen beschäftigten, konnte auf 13754 Datenbanktreffer (Ergebnis der Sucheingabe ohne arbeitsbezogene Begriffe am 13.06.2013) geschätzt werden. Die erste Studie wird 1952 beschrieben. Aus Abbildung 5 kann die Anzahl der jährlich veröffentlichten Arbeiten seit 1952 bis 2012 ersehen werden. Wie in Abschnitt 2.4.7 beschrieben, wurde mithilfe der Methode der kleinsten Quadrate eine nichtlineare Funktion erstellt, die den Anstieg der jährlichen Veröffentlichungszahlen beschreibt. Es lässt sich ein exponentieller Trend feststellen.

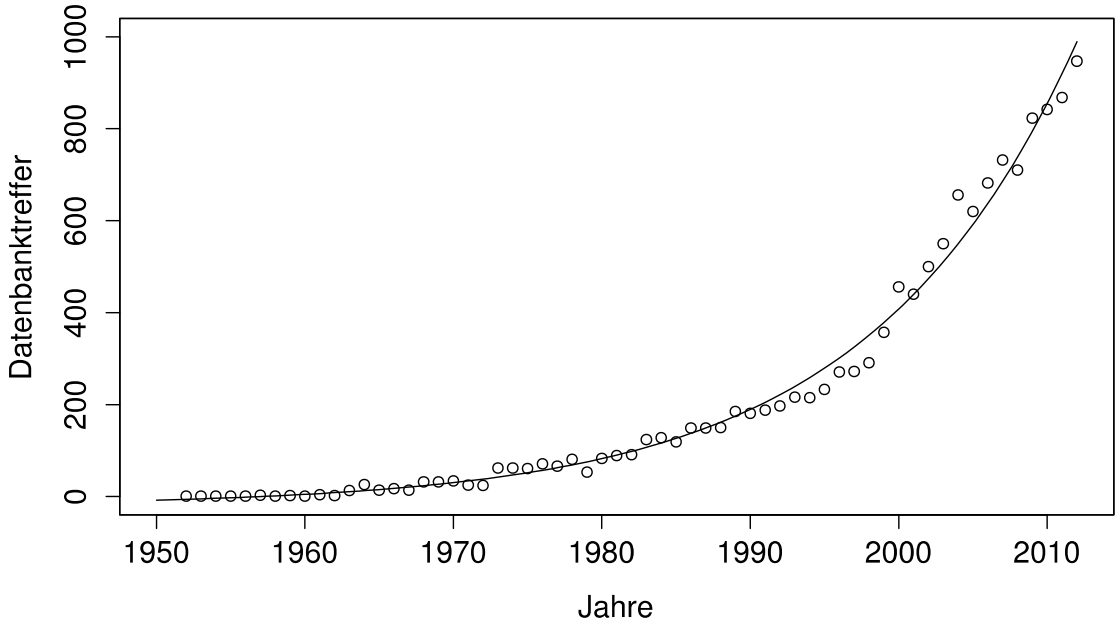


Abbildung 5: *Entwicklung der Studienveröffentlichungen über Erkrankungen mit chronischen Schmerzen.*

Nichtlinearer Fit : $Datenbanktreffer = e^{\frac{Jahre - 1951}{13{,}98}} \times \frac{1}{0{,}08} - 19{,}90$

Die 31 Publikationen, die tatsächlich über arbeitsbezogene Endpunkte berichteten, wurden vornehmlich ab den späten 90er Jahren veröffentlicht (Abbildung 6).

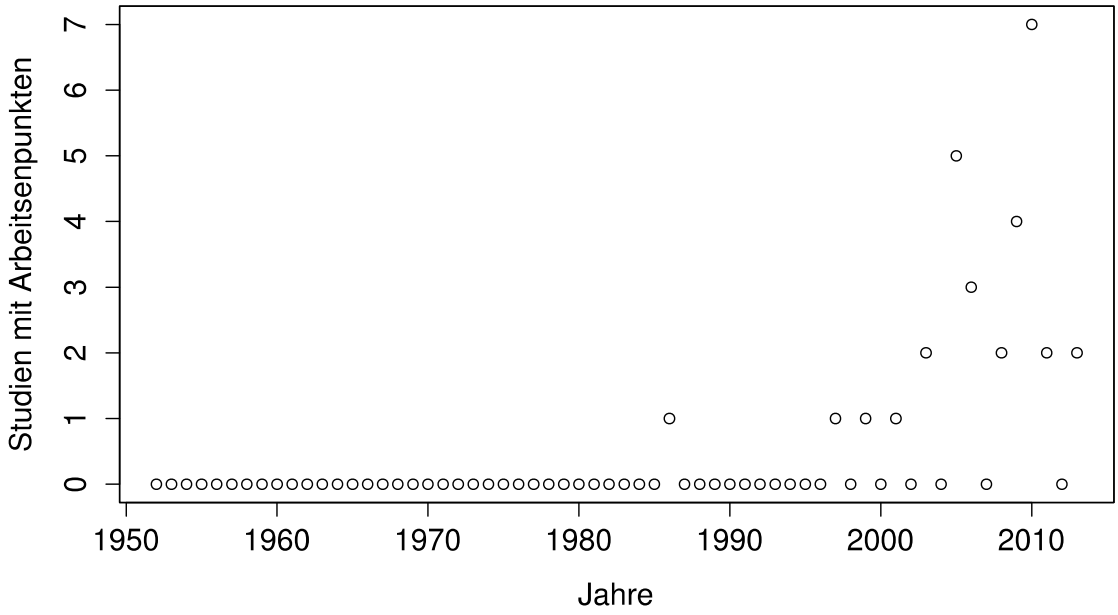


Abbildung 6: *Veröffentlichungen von Studien über chronische Schmerzerkrankungen, die arbeitsrelevante Endpunkte untersuchten.*

Diese 31 Publikationen machen also nur 0,23% der Grundgesamtheit aus. Dabei ist zu bemerken, dass der Anteil der Studien, die arbeitsbezogene Endpunkte berichteten, seit 1997 annähernd gleich blieb. Es wurde mit einem linearen Modell (Abschnitt 2.4.7) ein leicht positiver Trend konstruiert, der jedoch nicht statistisch signifikant ist (p=0,19) (Abbildung 7).

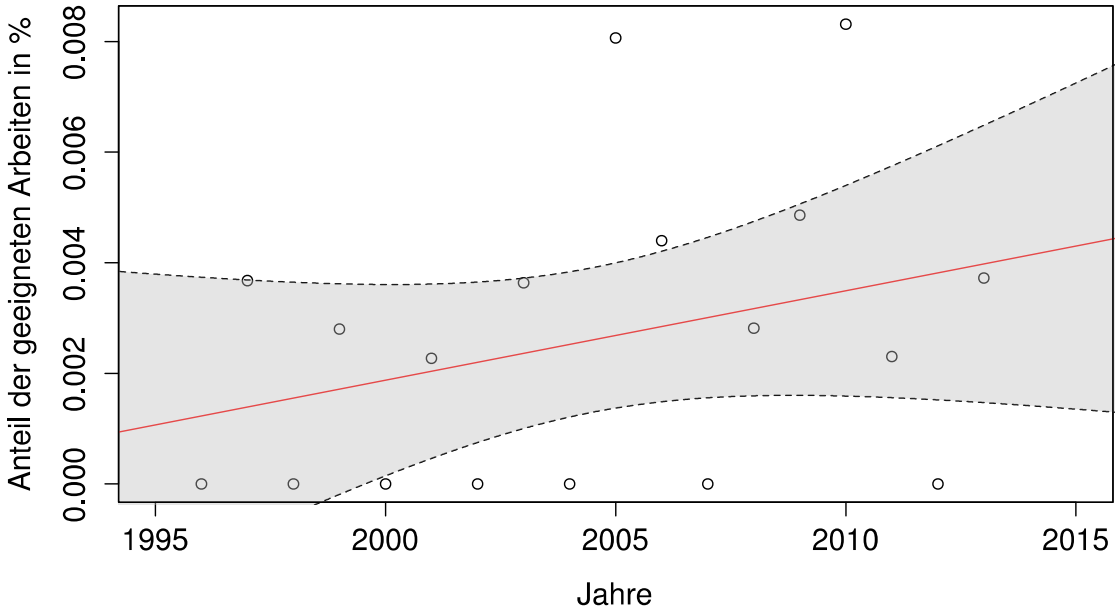


Abbildung 7: *Anteil der Studien, die arbeitsbezogene Endpunkte berichten, an der Gesamtanzahl der placebokontrollierten Studien über chronische Schmerzerkrankungen.*
Die gestrichelten Linien markieren das 95%-Konfidenzintervall

3.3 Wie wurde über arbeitsbezogene Endpunkte berichtet?

In den 31 in das systematische Review eingeschlossenen Arbeiten wurden 16 verschiedene arbeitsrelevante Endpunkte berichtet.

- Sieben Studien berichteten über die Beeinträchtigung am Arbeitsplatz als Unterpunkt des „Brief-Pain-Inventory“ (BPI).
  (Arnold et al. 2005, Chappell et al. 2009, Chappell et al. 2011, Markenson et al. 2005, Skljarevski et al. 2009, Skljarevski et al. 2010a, Skljarevski et al. 2010b)
- 13 Studien berichteten über verlorene Zeit am Arbeitsplatz.
  (Albert et al. 2013, Barkham et al. 2010, Bejarano et al. 2008, Bennett et al. 2003 & 2005, Egsmose et al. 1997, Kavanaugh et al. 2006, Kavanaugh et al. 2009a, Licciardone et al. 2003, Skljarevski et al. 2010a, Skljarevski et al. 2010b, Smolen et al. 2006, Straube et al. 2011, van der Heijde et al. 2006)
- Fünf untersuchten den Status der Beschäftigtenverhältnisse.
  (Bejarano et al. 2008, Carlsson und Sjölund 2001, Kavanaugh et al. 2006, Lehmann et al. 1986, Manchikanti et al. 2010)
- Vier befragten die Patienten über ihre Arbeitsfähigkeit.
  (Kavanaugh et al. 2006, Lehmann et al. 1986, Manchikanti et al. 2010, Smolen et al. 2006)
- Vier Studien enthielten Angaben zu den arbeitsrelevanten Fragen des „Short Form 36 Health Survey Questionnaire“ (SF-36).
  (Bennett et al. 2003 & 2005, Kavanaugh et al. 2006, Straube et al. 2011, van der Heijde et al. 2006)
- Drei erhoben die „Sheehan Disability Scale“ (SDS), welche unter anderem Patienten zu den Themen verpasste Arbeitszeit/Schulzeit und Behinderung bei diesen Tätigkeiten befragt.
  (Arnold et al. 2009, Chappell et al. 2008, Straube et al. 2011)

- Drei Studien benutzten den „Work Productivity and Activity Impairement Questionnaire“ (WPAI).
  (Baron et al. 2010, Skljarevski et al. 2010a, Skljarevski et al. 2010b)
- Drei Mal wurde eine visuelle Analogskala (VAS) verwendet, um selbst empfundene Arbeitsproduktivität zu erfassen.
  (Kavanaugh et al. 2006, Kavanaugh et al. 2013, van der Heijde et al. 2006)
- Drei Mal wurde über die arbeitsrelevanten Teilfragen des „Fibromyalgia Impact Questionnaire“ (FIQ) berichtet.
  (Bennett et al. 2003 & 2005, Bradley et al. 2010, Straube et al. 2011)
- Zwei Studien verwendeten eine „Work Instability Scale“ (WIS).
  (Barkham et al. 2010, Bejarano et al. 2008)
- Eine Studie erhob die „McGill Work Scale“. (Jarzem et al. 2005)
- Eine Studie berichtete über die Behinderung bei der Arbeit mithilfe des „Multidimensional Assessment of Fatigue“ (MAF). (Straube et al. 2011)
- Meireles et al. untersuchten 2010 den „Disabilities of the Arm, Shoulder and Hand Questionnaire“ (DASH).
- Ein Mal wurde die „Rheumatoid Arthritis Specific Work Productivity Survey“ (WPS-RA) berichtet. (Kavanaugh et al. 2009a)
- He et al. 2005 untersuchten Einschränkungen am Arbeitsplatz.
- Strand et al. 1999 untersuchten die Arbeitsproduktivität auf einer 0-100 Punkte Skala, die dem „Work Limitations Questionnaire of the National Opinion Research Center Survey“ entlehnt wurden.

Nicht nur in der schieren Anzahl der verschiedenen arbeitsbezogenen Endpunkte ließ sich eine deutliche Heterogenität feststellen. Die Studien unterschieden sich auch teils erheblich in Art der Erhebung beziehungsweise in der Beschreibung der einzelnen Endpunkte. So wurde der Endpunkt Arbeitsfehlzeiten auf insgesamt

sieben verschiedene Arten beschrieben: Vier Studien untersuchten die Fehlzeiten der letzten 7 Tage (Bennett et al. 2003 & 2005, Skljarevski et al. 2010a, Skljarevski et al. 2010b, Straube et al. 2011) , fünf untersuchten die gesamte Anzahl der Fehltage während der Dauer der Studie (Barkham et al. 2010, Bejarano et al. 2008, Egsmose et al. 1997, Smolen et al. 2006, van der Heijde et al. 2006). Je einmal wurden die Fehlzeiten im letzten Jahr (Albert et al. 2013), im letzten Monat (Kavanaugh et al. 2009a), in den letzten 6 Wochen (van der Heijde et al. 2006), in den letzten 4 Wochen (Licciardone et al. 2003) und schließlich während der letzten 4 Tage (Kavanaugh et al. 2006) berichtet.

Eine Aufstellung aller in der vorliegenden Dissertation ausgewerteten Veröffentlichungen mit den dazugehörigen, untersuchten arbeitsbezogenen Endpunkten findet sich in Tabelle 1.

3.4 Studienergebnisse

3.4.1 Gesamtüberblick

Um einen Eindruck zu gewinnen, wie sich arbeitsrelevante Parameter unter Behandlung entwickelten, wurden die Ergebnisse aller Studien ohne Berücksichtigung der Art ihrer Endpunkte betrachtet. Um eine mehrfache Einbeziehung von Patienten zu vermeiden, wurden die Arbeiten deren Patientenkollektive bereits von Bradley et al. 2010 erfasst wurden, aus dieser Betrachtung ausgeschlossen (Arnold et al 2005, Arnold et al. 2009, Chappell et al. 2008).

Für den Fall, dass mehrere Endpunkte untersucht worden sind, wurde entschieden, dass kontinuierliche validierte Instrumente (BPI, FIQ usw.) anderen Endpunkten vorzuziehen sind, um so die ohnehin große Heterogenität zwischen den verschiedenen Endpunkten so klein wie möglich zu halten. Wurden mehrere aktive Gruppen berichtet, so wurden die dazugehörigen Ergebnisse in einer gemeinsamen aktiven Gruppe gepoolt (Bradley et al. 2010, Jarzem et al. 2005, Skljarevski et al. 2009, Straube et al. 2011).

14 Studien (Bejarano et al. 2008, Bennett et al. 2003 & 2005, Bradley et al. 2010, Chappell et al. 2011, He et al. 2005, Kavanaugh et al. 2006, Kavanaugh et al. 2009a, Kavanaugh et al. 2013, Markenson et al. 2005, Meireles et al. 2010, Skljarevski et al. 2010a, Skljarevski et al. 2010b, Smolen et al. 2006, Strand et al. 1999) berichteten von einer statistisch signifikanten Verbesserung der Therapiegruppe gegenüber der Placebogruppe bezogen auf ihren Endpunkt. In acht Studien (Albert et al. 2013, Barkham et al. 2010, Chappell et al. 2009, Egsmose et al. 1997, Jarzem et al. 2005, Skljarevski et al. 2009, Straube et al. 2011, van der Heijde et al. 2006) wurde kein statistisch signifikanter Unterschied zwischen den Gruppen registriert. Drei Arbeiten (Baron et al. 2010, Lehmann et al. 1986, Licciardone et al. 2003) verzichteten aufgrund nicht erreichter Signifikanz auf das nähere Beschreiben ihrer Ergebnisse.

Carlsson und Sjölund 2001 und Manchikanti et al. 2010 führten für ihre Endpunkte keine statistischen Hypothesentests durch.

Im Vergleich der Therapie- und Placebogruppen berichteten durchschnittlich 56% (95%-KI: 35%; 76%) der Studien über ein statistisch signifikantes Ergebnis im Sinne einer Verbesserung der arbeitsbezogenen Endpunkte in der Therapiegruppe.

Die für den Gesamtüberblick relevanten Studien (Votecount) sind in Tabelle 3 zusammengefasst.

signifikante Verbesserung	**nicht signifikante Ergebnisse**
Bejarano et al. 2008	Albert et al. 2013
Bennett et al. 2003	Barkham et al. 2010
Bradley et al. 2010	Baron et al. 2010
Chappell et al. 2011	Chappell et al. 2009
He et al. 2005	Egsmose et al. 1997
Kavanaugh et al. 2006	Jarzem et al. 2005
Kavanaugh et al. 2009a	Lehmann et al. 1986
Kavanaugh et al. 2013	Licciardone et al. 2003
Meireles et al. 2010	Skljarevski et al. 2009
Markenson et al. 2005	Straube et al. 2011
Skljarevski et al. 2010a	van der Heijde et al. 2006
Skljarevski et al. 2010b	
Smolen et al. 2006	
Strand et al. 1999	

Signifikante Verschlechterungen wurden nicht berichtet.
Die Ergebnisse der aktiven Behandlungsgruppen wurden gepoolt.

Tabelle 3: *„Votecount“*
Ergebnisse der Endpunkte

3.4.1.1 Meta-Analyse der Durchschnittsdifferenzen aller arbeitsbezogenen Endpunkte

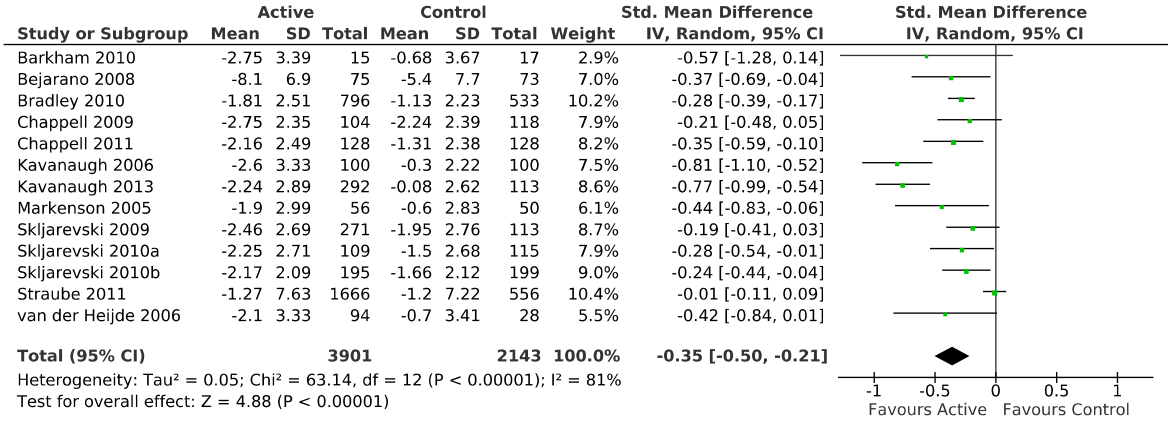
In 13 Studien wurde die Differenz bezüglich arbeitsbezogener Endpunkte zwischen Studienbeginn und -ende berichtet. In dieser Meta-Analyse wurden mehrere Endpunkte verglichen (Tabelle 4). Daher wurden diese mithilfe der zugehörigen Standardabweichung standardisiert und zusammengefasst.

	arbeitsbezogene Endpunkte
Barkham 2010	WIS
Bejarano 2008	WIS
Bradley 2010	FIQ
Chappell 2009	BPI-I
Chappell 2011	BPI-I
Kavanaugh 2006	VAS Produktivität
Kavanaugh 2013	VAS Produktivität
Markenson 2005	BPI-I
Skljarevski 2009	BPI-I
Skljarevski 2010a	BPI-I
Skljarevski 2010b	BPI-I
Straube 2011	FIQ
van der Heijde 2006	VAS Produktivität

BPI-I = Brief Pain Inventory: Interference with normal work;
FIQ = Fibromyalgia Impact Questionnaire;
VAS = Visual Analog Scale; WIS = Work Instability Scale

Tabelle 4: *Arbeitsbezogene Endpunkte der Studien (Mittelwertsdifferenzen)*

In diesen Studien wurden jeweils die Differenzen der untersuchten Endpunkte für die aktive Behandlungsgruppe (Active) und für die Kontrollgruppe (Control) gegeneinander verglichen. Da ein hohes Maß an Heterogenität, wegen der verschiedenen Behandlungen und Erkrankungen zu erwarten war und das Hauptanliegen der Arbeit ein genereller Überblick über arbeitsbezogene Endpunkte ist, wurde ein *random effects model* zur Meta-Analyse angewandt. Der so gepoolte standardisierte Behandlungseffekt beträgt -0,35 (95%-KI: -0,50; -0,21) und ist statistisch signifikant (Abbildung 8).



Study or Subgroup	Active Mean	Active SD	Active Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Barkham 2010	-2.75	3.39	15	-0.68	3.67	17	2.9%	-0.57 [-1.28, 0.14]
Bejarano 2008	-8.1	6.9	75	-5.4	7.7	73	7.0%	-0.37 [-0.69, -0.04]
Bradley 2010	-1.81	2.51	796	-1.13	2.23	533	10.2%	-0.28 [-0.39, -0.17]
Chappell 2009	-2.75	2.35	104	-2.24	2.39	118	7.9%	-0.21 [-0.48, 0.05]
Chappell 2011	-2.16	2.49	128	-1.31	2.38	128	8.2%	-0.35 [-0.59, -0.10]
Kavanaugh 2006	-2.6	3.33	100	-0.3	2.22	100	7.5%	-0.81 [-1.10, -0.52]
Kavanaugh 2013	-2.24	2.89	292	-0.08	2.62	113	8.6%	-0.77 [-0.99, -0.54]
Markenson 2005	-1.9	2.99	56	-0.6	2.83	50	6.1%	-0.44 [-0.83, -0.06]
Skljarevski 2009	-2.46	2.69	271	-1.95	2.76	113	8.7%	-0.19 [-0.41, 0.03]
Skljarevski 2010a	-2.25	2.71	109	-1.5	2.68	115	7.9%	-0.28 [-0.54, -0.01]
Skljarevski 2010b	-2.17	2.09	195	-1.66	2.12	199	9.0%	-0.24 [-0.44, -0.04]
Straube 2011	-1.27	7.63	1666	-1.2	7.22	556	10.4%	-0.01 [-0.11, 0.09]
van der Heijde 2006	-2.1	3.33	94	-0.7	3.41	28	5.5%	-0.42 [-0.84, 0.01]
Total (95% CI)			**3901**			**2143**	**100.0%**	**-0.35 [-0.50, -0.21]**

Heterogeneity: Tau² = 0.05; Chi² = 63.14, df = 12 (P < 0.00001); I² = 81%
Test for overall effect: Z = 4.88 (P < 0.00001)

Abbildung 8: *Meta-Analyse von arbeitsbezogenen Endpunkten über alle Studien hinweg, die Ergebnisse als Durchschnittsdifferenz berichteten. Durchschnittlicher Behandlungseffekt ausgedrückt als standardisierte Mittelwertsdifferenz.*

3.4.1.2 Heterogenitäts- und Sensitivitätsanalyse

Die Heterogenität der Meta-Analyse (Abschnitt 3.4.1.1) ist mit einem I^2-Wert von 81% als hoch anzusehen. Der Funnel-Plot in Abbildung 9 zeigt ein relativ homogenes Feld der Studien bei Bejarano et al. 2008, Bradley et 2010, Chappell et al. 2009, Chappell et al. 2011, Markenson et al. 2005, Skljarevski et al. 2009, Skljarevski et al. 2010a, Skljarevski et al. 2010b, van der Heijde et al. 2006. Außerhalb des Konfidenzintervalles sind die Arbeiten von Straube et al. 2011 mit einem sehr kleinen Standardfehler und kaum Effekt (-0,01) und von Kavanaugh et al. 2006 und Kavanaugh et al. 2013 mit einem eher durchschnittlichen Standardfehler, aber einem sehr großen Behandlungseffekt (-0,81 und -0,77) zu bemerken. Da der Hauptanteil der Studien auf der linken Hälfte des Funnel-Plots angesiedelt ist und bereits 3 erfasste Studien wegen mangelnder Signifikanz keine Daten veröffentlichten, lässt sich eine Veröffentlichungsverzerrung vermuten, aber aufgrund der geringen Bedeutung, die arbeitsbezogenen Endpunkten typischerweise bezüglich der Annahme oder Ablehnung solcher Publikationen zukommt, wohl kaum schlüssig folgern.

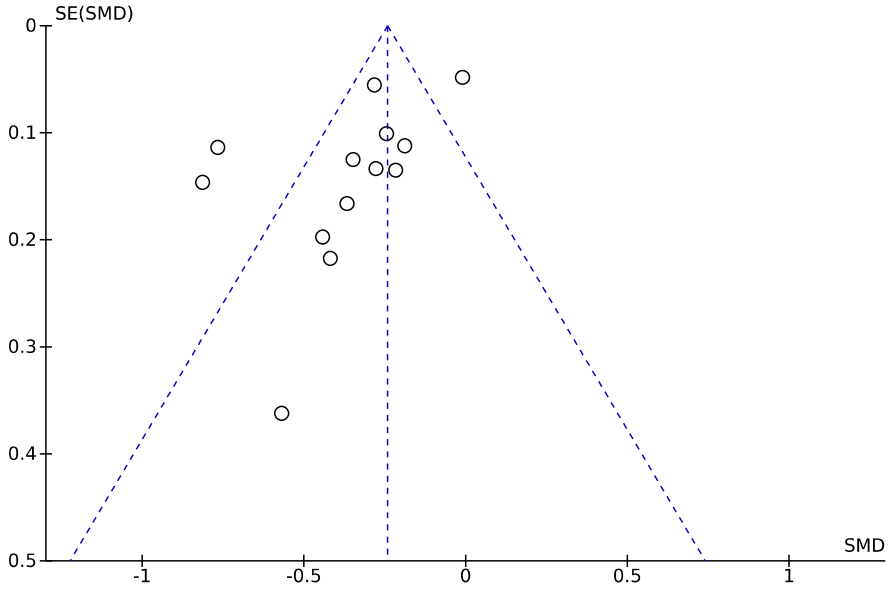


Abbildung 9: *Funnel-Plot der Gesamtüberblick-Meta-Analyse*

Die Arbeiten von Kavanaugh et al. in 2006 und 2013 unterschieden sich in ihren Ergebnissen recht stark von den Ergebnissen von Straube et al. 2011. Die drei genannten Studien wichen ebenfalls in ihren Ergebnissen von den restlichen Studien ab. Die Meta-Analyse von Straube et al. 2011 umfasste einen großen Patientenpool und machte mit 2222 Patienten von 6044 36,8% der gesamten Patienten aus. Gleichzeitig konnte diese über einen nur geringen Behandlungserfolg im Bereich der Verringerung der Beeinträchtigung bei der Arbeit bei Fibromyalgiepatienten berichten.

Um die vorliegende Heterogenität weiter zu ergründen, wurde ergänzend eine mit den standardisierten, inversen Varianzen der Studien gewichtete Regression durchgeführt, die einen möglichen Zusammenhang der Studienqualität mit dem Behandlungserfolg aufzeigen könnte. Obwohl die Studienqualität von der Publikation von Straube et al. 2011 nicht selbst bewertet wurde, da es sich um eine Meta-Analyse handelte, wurde ihr die Bewertung von 5/5 Punkten auf der OQS zugeteilt. Diese Vorgehensweise erschien vertretbar, da die Autoren explizit die Studienqualität der zu Grunde liegenden Studien berichteten. Es konnte ein

statistisch signifikanter Trend (Steigung 0,21; p = 0,03) gefunden werden, der darauf hinweist, dass qualitativ höher bewertete Arbeiten eventuell weniger ausgeprägte Veränderungen berichtet hatten (Abbildung 10).

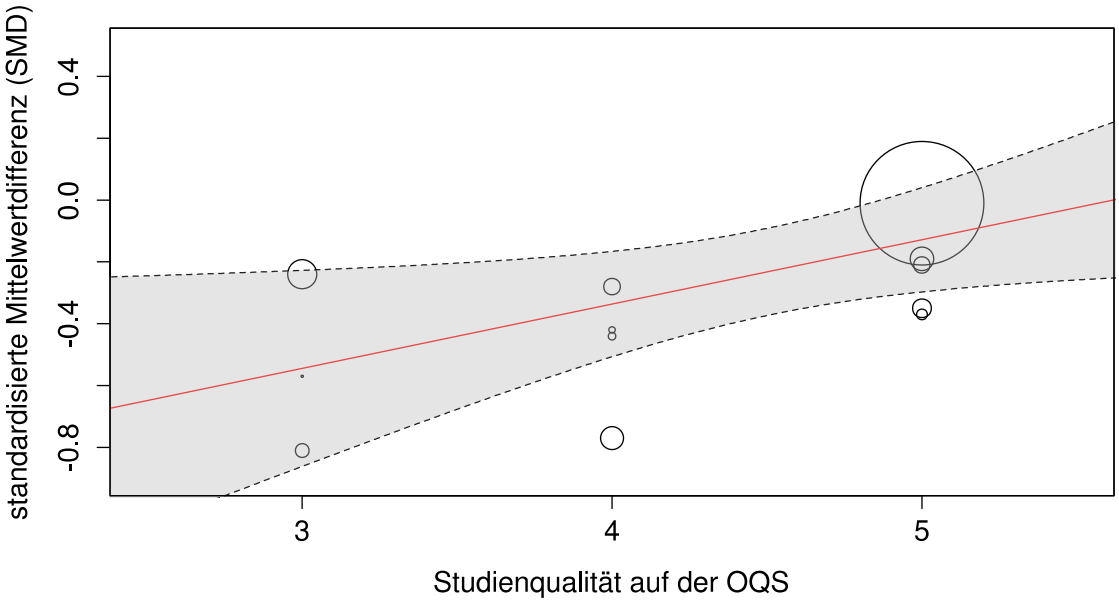


Abbildung 10: *Regression von Studienqualität auf der Oxford Quality Scale (OQS) mit dem standardisierten Studienergebnis, ausgedrückt als die standardisierte Mittelwertsdifferenz (SMD).Die Größe der Punkte repräsentiert die individuelle Gewichtung der einzelnen Studien, basierend auf ihrer inversen Varianz.*
Die gestrichelten Linien markieren das 95%-Konfidenzintervall.

Um einen Eindruck vom Einfluss der Heterogenität auf die Analyse in Abschnitt 3.4.1.1 zu erhalten, wurde eine weitere Meta-Analyse unter Ausschluss der Heterogenität verursachenden Studien (Kavanaugh et al. 2006, Kavanaugh et al. 2013 und Straube et al. 2011) mit einem *fixed effect model* erstellt. Der erhaltene mit -0,28 (95%-KI: -0,35; -0,21) geschätzte Behandlungseffekt ist mit einem p-Wert kleiner als 0.00001 statistisch signifikant. Das Ergebnis liegt nahe an der vorherigen Analyse mit einem *random effects model* (-0,35). Dies spricht dafür, dass in allen Studien ähnliche Effekte gemessen wurden. Das *random effects model* bildet in diesem Zusammenhang jedoch mit seinem deutlich weiteren Konfidenzintervall und der Verwendung aller verfügbaren Informationen ein konservativeres und verlässlicheres Ergebnis ab.

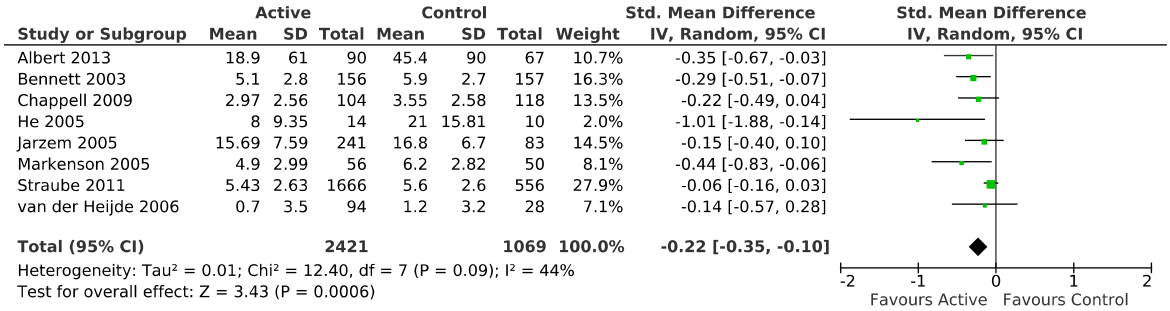
3.4.1.3 Meta-Analyse aller arbeitsbezogenen Endpunkte (Daten zu Studienende)

Zu ähnlichen Ergebnissen wie die Meta-Analyse in Abschnitt 3.4.1.1, kam auch die Meta-Analyse der Studien (Abbildung 11), die Baselinedaten und Daten zu Studienende veröffentlichten (Albert et al. 2013, Bennett et al. 2003, Chappell et al. 2009, He et al. 2005, Jarzem et al. 2005, Markenson et al. 2005, Straube et al. 2011, van der Heijde et al. 2006). Bei diesen Studien wurden die zuletzt erfassten Werte der verschiedenen Endpunkte (Tabelle 5) in einem *random effects model* zusammengefasst. Die Studie von Meireles et al. 2010 wurde dabei ausgeschlossen, da sich die Patienten der Behandlungsarme (aktive Behandlungsgruppe und Kontrollgruppe) bereits zum Zeitpunkt des Studienbeginns stark unterschieden (t-Test, p = 0,05).

	arbeitsbezogene Endpunkte
Albert 2013	Arbeitszeitverluste
Bennett 2003	FIQ
Chappell 2009	BPI-I
He 2005	Einschränkungen am Arbeitsplatz
Jarzem 2005	McGill Work Scale
Markenson 2005	BPI-I
Straube 2011	FIQ
van der Heijde 2006	Arbeitszeitverluste

BPI-I = Brief Pain Inventory: Interference with normal work;
FIQ = Fibromyalgia Impact Questionnaire

Tabelle 5: *Arbeitsbezogene Endpunkte der Studien (zu Studienende)*



Study or Subgroup	Active Mean	Active SD	Active Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Albert 2013	18.9	61	90	45.4	90	67	10.7%	-0.35 [-0.67, -0.03]
Bennett 2003	5.1	2.8	156	5.9	2.7	157	16.3%	-0.29 [-0.51, -0.07]
Chappell 2009	2.97	2.56	104	3.55	2.58	118	13.5%	-0.22 [-0.49, 0.04]
He 2005	8	9.35	14	21	15.81	10	2.0%	-1.01 [-1.88, -0.14]
Jarzem 2005	15.69	7.59	241	16.8	6.7	83	14.5%	-0.15 [-0.40, 0.10]
Markenson 2005	4.9	2.99	56	6.2	2.82	50	8.1%	-0.44 [-0.83, -0.06]
Straube 2011	5.43	2.63	1666	5.6	2.6	556	27.9%	-0.06 [-0.16, 0.03]
van der Heijde 2006	0.7	3.5	94	1.2	3.2	28	7.1%	-0.14 [-0.57, 0.28]
Total (95% CI)			**2421**			**1069**	**100.0%**	**-0.22 [-0.35, -0.10]**

Heterogeneity: Tau² = 0.01; Chi² = 12.40, df = 7 (P = 0.09); I² = 44%
Test for overall effect: Z = 3.43 (P = 0.0006)

Abbildung 11: *Meta-Analyse von arbeitsbezogenen Endpunkten über alle Studien hinweg, die Daten zum Studienende berichteten. Durchschnittlicher Behandlungseffekt ausgedrückt als standardisierte Mittelwertsdifferenz.*

Der standardisierte mittlere Unterschied zwischen den Behandlungsgruppen betrug -0,22 (95%-KI: -0,35; -0,10) und ist mit p=0,0006 statistisch signifikant.

3.4.2 Meta-Analyse zu „Brief Pain Inventory: Interference with normal work“

Das in den Studien am homogensten untersuchte arbeitsbezogene Messinstrument war die Teilfrage des „Brief Pain Inventory“: Beeinträchtigung bei normaler Arbeit. Insgesamt sieben Studien berichteten über die Differenz zwischen Studienbeginn und -ende (Arnold et al. 2005, Chappell et al. 2009, Chappell et al. 2011, Markenson et al. 2005, Skljarevski et al. 2009, Skljarevski et al. 2010a, Skljarevski et al. 2010b). Da hier alle Studien den gleichen Endpunkt untersuchten, war eine Standardisierung nicht notwendig. Das Ergebnis kann direkt in einer fassbaren Größe (BPI-Beeinträchtigungspunkte) ausgedrückt werden (Abbildung 12).

Study or Subgroup	Active Mean	Active SD	Active Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
Arnold 2005	-2.52	2.51	230	-1.2	2.5	118	16.4%	-1.32 [-1.88, -0.76]
Chappell 2009	-2.75	2.35	104	-2.24	2.39	118	13.6%	-0.51 [-1.13, 0.11]
Chappell 2011	-2.16	2.49	128	-1.31	2.38	128	14.7%	-0.85 [-1.45, -0.25]
Markenson 2005	-1.9	2.99	56	-0.6	2.83	50	5.0%	-1.30 [-2.41, -0.19]
Skljarevski 2009	-2.46	2.69	271	-1.95	2.76	113	14.5%	-0.51 [-1.11, 0.09]
Skljarevski 2010a	-2.25	2.71	109	-1.5	2.68	115	11.1%	-0.75 [-1.46, -0.04]
Skljarevski 2010b	-2.17	2.09	195	-1.66	2.12	199	24.7%	-0.51 [-0.93, -0.09]
Total (95% CI)			**1093**			**841**	**100.0%**	**-0.76 [-1.02, -0.50]**

Heterogeneity: Tau² = 0.02; Chi² = 7.56, df = 6 (P = 0.27); I² = 21%
Test for overall effect: Z = 5.79 (P < 0.00001)

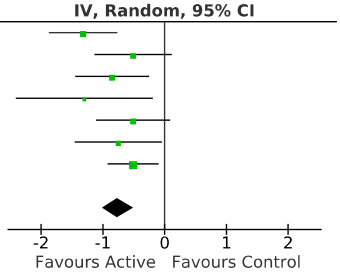


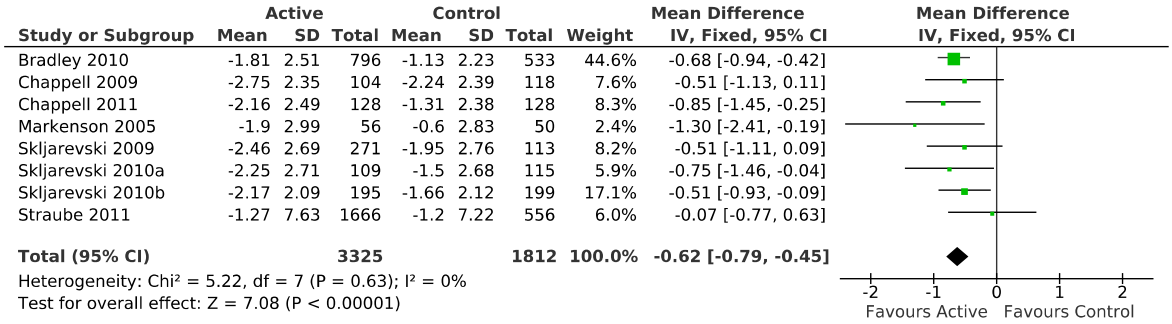
Abbildung 12: *Meta-Analyse aller Studien, die den BPI Unterpunkt „Interference with normal work“ berichteten. Der Behandlungseffekt wird als Veränderung auf einer Skala von 0-10 Punkten ausgedrückt.*

Die Heterogenität ist mit einem I^2-Wert von 21% als niedrig einzustufen.

Es ergab sich ein statistisch signifikanter Behandlungseffekt von -0,76 BPI-Punkten (95%-KI: -1,02; -0,50).

3.4.3 Meta-Analyse zu „Interference with work“ (BPI/FIQ)

Die Unterpunkte, die sich mit Beeinträchtigungen der Arbeitsfähigkeit beschäftigen, sind für das BPI und den FIQ sehr ähnlich formuliert. Daher ist es möglich, diese mit den Studien aus der Meta-Analyse in Abschnitt 3.4.2 zu kombinieren, ohne eine Standardisierung durchführen zu müssen. Bradley et al. 2010 schließen unter anderen bereits die Studie von Arnold et al. 2005 mit ein, so dass für die hier vorgestellten Betrachtungen statt der Primärpublikation die Meta-Analyse von Bradley et al. 2010 verwendet wurde. Bennett et al. 2003 & 2005 untersuchten ebenfalls den FIQ, berichteten jedoch nicht die Mittelwertsdifferenzen und wurden deshalb nicht berücksichtigt. Weiterhin wurde die Arbeit von Straube et al. 2011 in die Meta-Analyse aufgenommen (Abbildung 13).



Study or Subgroup	Active Mean	Active SD	Active Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Fixed, 95% CI
Bradley 2010	-1.81	2.51	796	-1.13	2.23	533	44.6%	-0.68 [-0.94, -0.42]
Chappell 2009	-2.75	2.35	104	-2.24	2.39	118	7.6%	-0.51 [-1.13, 0.11]
Chappell 2011	-2.16	2.49	128	-1.31	2.38	128	8.3%	-0.85 [-1.45, -0.25]
Markenson 2005	-1.9	2.99	56	-0.6	2.83	50	2.4%	-1.30 [-2.41, -0.19]
Skljarevski 2009	-2.46	2.69	271	-1.95	2.76	113	8.2%	-0.51 [-1.11, 0.09]
Skljarevski 2010a	-2.25	2.71	109	-1.5	2.68	115	5.9%	-0.75 [-1.46, -0.04]
Skljarevski 2010b	-2.17	2.09	195	-1.66	2.12	199	17.1%	-0.51 [-0.93, -0.09]
Straube 2011	-1.27	7.63	1666	-1.2	7.22	556	6.0%	-0.07 [-0.77, 0.63]
Total (95% CI)			**3325**			**1812**	**100.0%**	**-0.62 [-0.79, -0.45]**

Heterogeneity: Chi² = 5.22, df = 7 (P = 0.63); I² = 0%
Test for overall effect: Z = 7.08 (P < 0.00001)

Abbildung 13: *Meta-Analyse aller Studien, die den Unterpunkt „Interference with work“ berichteten (BPI/FIQ). Der Behandlungseffekt wird als Veränderung auf einer Skala von 0-10 Punkten ausgedrückt.*

Hier konnte keine Heterogenität festgestellt werden. Die durchschnittliche Mittelwertsdifferenz betrug -0,62 (95%-KI: -0,79; -0,45). Das statistisch signifikante Ergebnis ähnelte darin der vorherigen Meta-Analyse (Abschnitt 3.4.2).

3.4.3.1 Vergleich der Ergebnisse auf der BPI/FIQ-Skala und der standardisierten Ergebnisse

Um die Ergebnisse der ersten Meta-Analyse des Gesamtüberblickes (Abschnitt 3.4.1.1) mit der Analyse der Beeinträchtigung bei der Arbeit des BPI und FIQ zu vergleichen (Abbildung 13), wurde die SMD aus Abschnitt 3.4.1.1 mithilfe der Standardabweichung des BPI/FIQ-Unterpunktes ausgedrückt. Von den 7 Studien, die den BPI oder das FIQ berichteten, dokumentierten nur Bradley et al. 2010 die Baseline-Standardabweichungen der Beeinträchtigung bei der Arbeit. Wie von Higgins und Green 2011 beschrieben, wurden die Standardabweichungen der verschiedenen Behandlungsgruppen aus der Arbeit von Bradley et al. 2010 gepoolt. Die daraus folgende Standardabweichung betrug 2,09. Dies stellte den Umrechnungsfaktor für die Darstellung der Ergebnisse aus Abschnitt 3.4.1.1 auf der 0-10 Punkte-Skala von BIP und FIQ dar. Es ist zu beachten, dass dieses Verfahren Einschränkungen unterliegt, die in Abschnitt 4.2.1.2 näher beschrieben sind.

Die so errechnete Mittelwertsdifferenz der Gesamtüberblick-Meta-Analyse (Abschnitt 3.4.1.1) betrug -0,73 Beeinträchtigungspunkte (95%-KI: -1,00; -0,44).

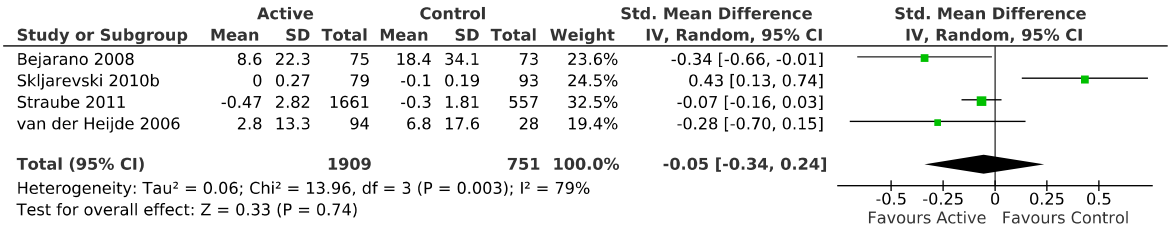
Das Ergebnis der standardisierten Mittelwertsdifferenzen war damit den direkt gepoolten Ergebnissen in Abschnitt 3.4.3 ähnlich.

3.4.4 Meta-Analyse zu Arbeitsfehlzeiten

Die Berichterstattung bezüglich Arbeitsfehlzeiten war in den untersuchten Studien heterogen. Von den 13 Studien, die Arbeitsfehlzeiten untersuchten, listeten nur sechs verwertbare Daten auf (Albert et al. 2013, Bejarano et al. 2008, Bennett et al. 2003, Skljarevski et al. 2010b, Straube et al. 2011, van der Heijde et al. 2006). Drei berichteten keinerlei genauere Ergebnisse (Barkham et al. 2010, Licciardone et al. 2003, Skljarevski et al. 2010a). Die Verwendung der Studien von Egsmose et al. 1997, Kavanaugh et al. 2006 und Kavanaugh et al. 2009a scheiterte an mangelnder Berichterstattung, da die Autoren dieser Studien darauf verzichteten Standardabweichung oder Standardfehler zu dokumentieren. Smolen et al. 2006 waren die einzigen, die diesen Endpunkt kategorisch berichteten und konnten daher nicht mit anderen Studien verglichen werden.
Wie bereits in Abschnitt 3.1 beschrieben, wurden in den betrachteten Studien viele verschiedene Formen von Arbeitsfehlzeiten erfasst, so dass die Ergebnisse für die Verwendung in den folgenden Meta-Analysen standardisiert werden mussten.

3.4.4.1 Meta-Analyse zu Arbeitsfehlzeiten berichtet als Durchschnittsdifferenzen

In diese Meta-Analyse flossen die Daten der Studien ein, die Durchschnitts-differenz von Studienbeginn und -ende berichteten (Skljarevski et al. 2010b, Straube et al. 2011) sowie die Daten der Studien, die Arbeitszeitverluste während der Studie registrierten (Bejarano et al. 2008, van der Heijde et al. 2006) (Abbildung 14).



Study or Subgroup	Active Mean	Active SD	Active Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bejarano 2008	8.6	22.3	75	18.4	34.1	73	23.6%	-0.34 [-0.66, -0.01]
Skljarevski 2010b	0	0.27	79	-0.1	0.19	93	24.5%	0.43 [0.13, 0.74]
Straube 2011	-0.47	2.82	1661	-0.3	1.81	557	32.5%	-0.07 [-0.16, 0.03]
van der Heijde 2006	2.8	13.3	94	6.8	17.6	28	19.4%	-0.28 [-0.70, 0.15]
Total (95% CI)			**1909**			**751**	**100.0%**	**-0.05 [-0.34, 0.24]**

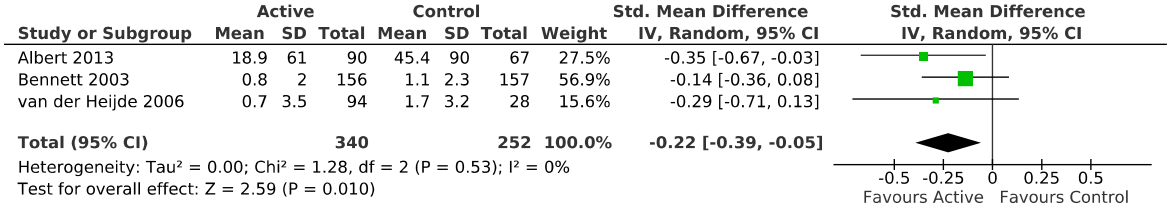
Heterogeneity: Tau² = 0.06; Chi² = 13.96, df = 3 (P = 0.003); I² = 79%
Test for overall effect: Z = 0.33 (P = 0.74)

Abbildung 14: *Meta-Analyse der Studien, die Arbeitsfehlzeiten als Durchschnittsdifferenz berichteten. Behandlungseffekt ausgedrückt als standardisierte Mittelwertsdifferenz.*

Es konnte kein statistisch signifikanter Behandlungseffekt festgestellt werden.

3.4.4.2 Meta-Analyse zu Arbeitsfehlzeiten berichtet zu Studienende

Eine Meta-Analyse der Studien, die Werte zu Studienende zur Verfügung stellten (Albert et al. 2013, Bennett et al. 2003, van der Heijde et al. 2006), zeigte einen deutlichen Unterschied zwischen aktiver und Placebogruppe (Abbildung 15).



Study or Subgroup	Active Mean	Active SD	Active Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Albert 2013	18.9	61	90	45.4	90	67	27.5%	-0.35 [-0.67, -0.03]
Bennett 2003	0.8	2	156	1.1	2.3	157	56.9%	-0.14 [-0.36, 0.08]
van der Heijde 2006	0.7	3.5	94	1.7	3.2	28	15.6%	-0.29 [-0.71, 0.13]
Total (95% CI)			**340**			**252**	**100.0%**	**-0.22 [-0.39, -0.05]**

Heterogeneity: Tau² = 0.00; Chi² = 1.28, df = 2 (P = 0.53); I² = 0%
Test for overall effect: Z = 2.59 (P = 0.010)

Abbildung 15: *Meta-Analyse der Studien, die die Arbeitsfehlzeiten zum Zeitpunkt des Studienendes berichteten. Behandlungseffekt ausgedrückt als standardisierte Mittelwertsdifferenz.*

Es wurde ein standardisierter Behandlungseffekt von –0,22 (95%-KI: -0,39; -0,05) errechnet, der mit p=0,01 statistisch signifikant ist.

3.4.4.3 Vergleich der Ergebnisse aus den Abschnitten 3.4.4.1 und 3.4.4.2

Bei der Beurteilung der Ergebnisse dieser beiden Meta-Analysen müssen die Form der Berichterstattung und eine eventuelle Heterogenität der zu Grunde liegenden Daten beachtet werden. Die Daten der ersteren Analyse beziehen bereits vor der Studie bestehende Unterschiede zwischen den Studienarmen ein, während die zweite sich vollständig auf die Randomisierung zur Bereinigung eventueller Variabilitäten verlässt.

Bezogen auf Arbeitsfehlzeiten verzeichnete die Studie von Skljarevski et al. 2010b am Studienende eine signifikante Verschlechterung der aktiven Gruppe gegenüber der Kontrollgruppe. So unterschied sich der geschätzte Behandlungseffekt dieser Studie deutlich von den der anderen Studien aus der Analyse in Abschnitt 3.4.4.1.

Aufgrund der uneinheitlichen und teilweise ungenauen Berichterstattung sowie der großen Heterogenität in der Meta-Analyse in Abschnitt 3.4.4.1 ist die Frage, ob sich die Behandlungen in den untersuchten Studien positiv auf die Arbeitsfehlzeiten der Patienten auswirkten, nicht abschließend zu beantworten.

3.4.5 Arbeitsproduktivität

Sieben Studien berichteten über den Endpunkt der Arbeitsproduktivität.

Davon konnte aus 4 Studien (Kavanaugh et al. 2006, Kavanaugh et al. 2013, Skljarevski et al. 2010b, van der Heijde et al. 2006) die Differenz zwischen Studienbeginn und -ende extrahiert werden, während die restlichen drei Arbeiten keine verwertbaren Daten berichteten (Baron et al. 2010, Kavanaugh et al. 2009a, Skljarevski et al. 2010a).

3.4.5.1 Meta-Analyse zu erfasster Arbeitsproduktivität

Von den vier verwertbaren Studien verwendeten nur Skljarevski et al. 2010b keine visuelle Analogskala, um Arbeitsproduktivität zu messen, sondern berichteten über einen Unterpunkt der „Work Productivity and Activity Impairment Questionnaire“ (WPAI). Aufgrund der unterschiedlichen Beschaffenheit der Fragestellungen wurde wieder mit standardisierten Mittelwertsdifferenzen gearbeitet (Abbildung 16).

	Active			Control				Std. Mean Difference
Study or Subgroup	**Mean**	**SD**	**Total**	**Mean**	**SD**	**Total**	**Weight**	**IV, Random, 95% CI**
Kavanaugh 2006	-2.6	3.33	100	-0.3	2.22	100	25.7%	-0.81 [-1.10, -0.52]
Kavanaugh 2013	-2.24	2.89	292	-0.08	2.62	113	27.7%	-0.77 [-0.99, -0.54]
Skljarevski 2010b	-0.18	0.26	77	-0.16	0.28	89	25.2%	-0.07 [-0.38, 0.23]
van der Heijde 2006	-2.1	3.33	94	-0.7	3.41	28	21.3%	-0.42 [-0.84, 0.01]
Total (95% CI)			**563**			**330**	**100.0%**	**-0.53 [-0.88, -0.18]**

Heterogeneity: Tau² = 0.10; Chi² = 16.23, df = 3 (P = 0.001); I² = 82%
Test for overall effect: Z = 2.97 (P = 0.003)

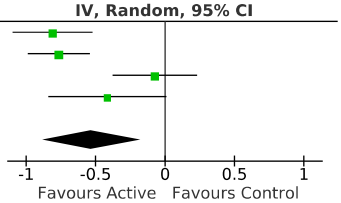


Abbildung 16: *Meta-Analyse der Studien, die Arbeitsproduktivität als Durchschnittsdifferenz berichteten. Behandlungseffekt ausgedrückt als standardisierte Mittelwertsdifferenz.*

Mit 82% ist I^2 als hoch einzustufen. Der statistisch signifikante standardisierte Behandlungseffekt betrugt -0,53 (95%-KI: -0,88; -0,18).

3.4.5.2 Meta-Analyse zu erfasster Arbeitsproduktivität nur der Visual-Analog-Scale-Studien

Um eine Auswertung ohne Standardisierung zu ermöglichen, wurden nur die Studien betrachtet, die ihre Ergebnisse mithilfe von visuellen Analogskalen verglichen. Dabei musste die Studie von Skljarevski et al. 2010b ausgeschlossen werden, da diese Ergebnisse der WPAI berichtete.

In der Meta-Analyse mit einem *fixed effect model* zeigte sich keinerlei messbare Heterogenität (Abbildung 17).

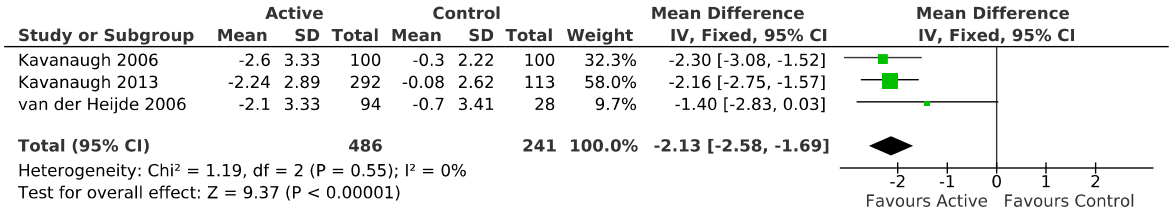


Abbildung 17: *Meta-Analyse der Studien, die Arbeitsproduktivität auf einer visuellen Analogskala erfassten und als Durchschnittsdifferenz berichteten. Der Behandlungseffekt wurde als Unterschied in den Bewertungen auf einer 0-10 visuellen Analogskala berichtet.*

Durchschnittlich hatten sich die Patienten um 2,13 von 10 Produktivitätseinheiten (95%-KI: 1,69, 2,58) verbessert.

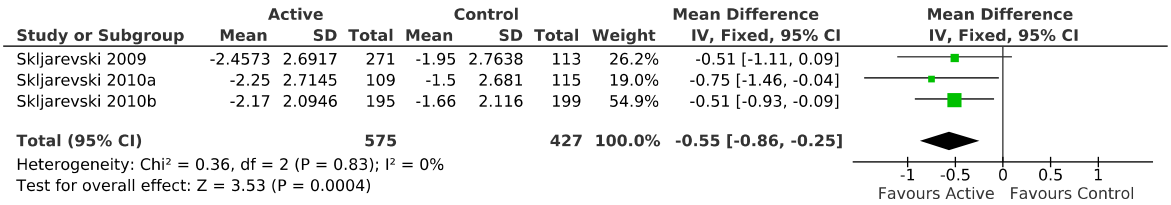
3.4.5.3 Vergleich der Ergebnisse auf der VAS-Skala und der standardisierten Ergebnisse

Alle in Abschnitt 3.4.5.2 verwendeten Studien berichteten die Standard-abweichungen ihres Endpunktes zu Studienbeginn. Der daraus gepoolte Wert (3,07) diente, gemäß Higgins und Green 2011 (Abschnitt 4.2.1.2), als Umrechnungsfaktor für die standardisierten Mittelwertsdifferenzen. So ergab sich für die in Abschnitt 3.4.5.1 durchgeführte Meta-Analyse, bei der die Studie von Skljarevski et al. 2010b mit einfloss, eine Verbesserung von 1,63 auf der 0-10 VAS (95%-KI: 0,55; 2,70).

Wurde der Umrechnungsfaktor auf die Ergebnisse des Gesamtüberblickes aus Abschnitt 3.4.1.1 angewendet, so ergab sich eine Produktivitätsverbesserung von 1,07 (95%-KI: 0,64; 1,54) auf der VAS.

3.4.6 Meta-Analyse zu „Brief Pain Inventory: Interference with normal work“ bei chronischen Rückenschmerzen

Die Gruppe um Skljarevski et al. führte drei Studien, publiziert in den Jahren 2009 und 2010, durch, um den Effekt von Duloxetin auf chronische Rückenschmerzen zu untersuchen. Da nicht überlappende Patientenpopulationen untersucht wurden, konnten die Ergebnisse dieser Studien in einer Meta-Analyse zusammengeführt werden. Bei der Teilfrage der Beeinträchtigung bei normaler Arbeit des BPI konnte nach dem Poolen der drei Arbeiten ein statistisch signifikanter Behandlungserfolg festgestellt werden (Abbildung 18).



	Active			Control				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
Skljarevski 2009	-2.4573	2.6917	271	-1.95	2.7638	113	26.2%	-0.51 [-1.11, 0.09]
Skljarevski 2010a	-2.25	2.7145	109	-1.5	2.681	115	19.0%	-0.75 [-1.46, -0.04]
Skljarevski 2010b	-2.17	2.0946	195	-1.66	2.116	199	54.9%	-0.51 [-0.93, -0.09]
Total (95% CI)			**575**			**427**	**100.0%**	**-0.55 [-0.86, -0.25]**

Heterogeneity: Chi² = 0.36, df = 2 (P = 0.83); I² = 0%
Test for overall effect: Z = 3.53 (P = 0.0004)

Abbildung 18: *Meta-Analyse der Arbeiten, die den BPI Unterpunkt „Interference with normal work“ (BPI-I) in Studien über chronische Rückenschmerzen berichteten. Der Behandlungseffekt wird als Veränderung auf einer Skala von 0-10 Punkten ausgedrückt.*

Der allgemeine Behandlungseffekt gegenüber Placebo konnte auf -0,55 auf einer Skala von 10 BPI-Punkten geschätzt werden (95%-KI: -0,86; -0,25). Da alle Studien von der gleichen Arbeitsgruppe durchgeführt wurden, überrascht das Fehlen von Heterogenität nicht. Die Anwendung des *fixed effect models* war daher gerechtfertigt.

3.4.7 Arbeitsfähigkeit

Zur Abschätzung der Auswirkungen der Behandlungen je nach Ausgangssituation ist es notwendig, die Entwicklung von noch arbeitenden und bereits arbeitsunfähigen Patienten getrennt zu untersuchen. Dazu wurden, wie in Abschnitt 2.4.6 ausgeführt, für die einzelnen Studien Wahrscheinlichkeits-quotienten und Wahrscheinlichkeitsdifferenzen berechnet und in Meta-Analysen gepoolt.

3.4.7.1 Meta-Analyse zu der Vermeidung von Arbeitsunfähigkeit

Aus den 5 Studien, die den Beschäftigungsstatus der Patienten untersuchten, konnten nur bei Carlsson und Sjölund 2001 und Smolen et al. 2006 zuverlässige Daten extrahiert werden, um einen Wahrscheinlichkeitsquotienten zu berechnen.

Von den fünf Studien, die den Beschäftigungsstatus der Patienten untersuchten, konnten drei nicht in dieser Meta-Analyse verwendet werden. Lehmann et al. 1986 berichteten über die Rückkehr in ein Arbeitsverhältnis ihrer Patienten, unterschieden aber nicht zwischen aktiver Behandlungsgruppe und Kontrollgruppe. Manchikanti et al. 2010 berichteten über Beschäftigungsstatus vor und nach der Studie, es konnte jedoch nicht zugeordnet werden, welcher Status sich wie verändert hatte. Kavanaugh et al. 2006 berichtete dagegen nicht über die Entwicklung von noch arbeitsfähigen Probanden, sondern nur über die Rehabilitation von bereits arbeitsunfähigen Patienten.
Die zwei Studien, die verwertbare Daten berichteten (Carlsson und Sjölund 2001, Smolen et al. 2006), schätzten den Wahrscheinlichkeitsquotienten (RR) arbeitsunfähig zu werden recht ähnlich ein. Letztere Studie dominiert mit ca 97% der Patienten die erstellte Meta-Analyse. Es war keine relevante Heterogenität zwischen den Studien nachweisbar (Abbildung 19).

Study or Subgroup	Active Events	Active Total	Control Events	Control Total	Weight	Risk Ratio IV, Random, 95% CI
Carlsson & Sjölund 2001	0	13	1	6	2.6%	0.17 [0.01, 3.59]
Smolen 2006	34	406	22	159	97.4%	0.61 [0.37, 1.00]
Total (95% CI)		**419**		**165**	**100.0%**	**0.59 [0.36, 0.96]**
Total events	34		23			

Heterogeneity: Tau² = 0.00; Chi² = 0.66, df = 1 (P = 0.42); I² = 0%
Test for overall effect: Z = 2.11 (P = 0.03)

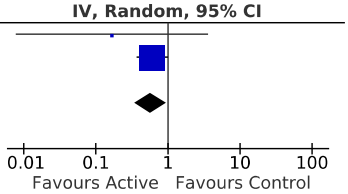


Abbildung 19: *Meta-Analyse der Wahrscheinlichkeitsquotienten, war man zu Beginn der Studie arbeitsfähig, zu Studienende arbeitsunfähig zu sein.*

Der errechnete gemeinsame RR, zu Studienende arbeitsunfähig zu sein, wenn man zu Beginn der Behandlung arbeitsfähig war, betrug 0,59. Die Behandlungen senkten damit das Risiko, arbeitsunfähig zu werden, um 41% (95%-KI: 4%; 64%). Das Ergebnis konnte mit p=0,03 statistische Signifikanz vorweisen.

Eine weiteren Analyse der Wahrscheinlichkeitsdifferenzen ergab zwischen den Behandlungsgruppen einen Unterschied von 6% (95%-KI: 0%; 12%).

3.4.7.2 Meta-Analyse zu der Wiederherstellung von Arbeitsfähigkeit

Bei der Analyse der Wiederherstellung der Arbeitsfähigkeit konnte, zusätzlich zu den in 3.4.7.1 verwendeten Studien, die Studie der Gruppe um Kavanaugh et al. 2006 mit einbezogen werden (Abbildung 20), da die Autoren über die rehabilitierten Patienten berichteten.

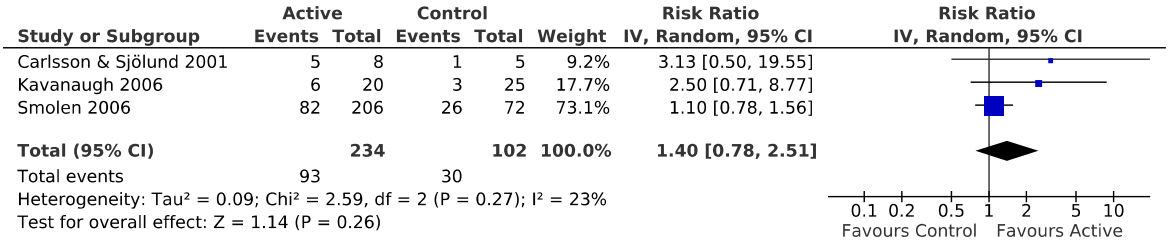


Abbildung 20: *Meta-Analyse der Wahrscheinlichkeitsquotienten, war man zu Beginn der Studie arbeitsunfähig, zu Studienende wieder arbeitsfähig zu sein.*

Mit einem I^2-Wert von 23% ist der Effekt von Heterogenität auf die Analyse noch als gering einzustufen. Die auf ein *random effects model* basierende Meta-Analyse zeigt ein nicht signifikantes (p=0,26) relatives Risiko von 1,40 (95%-KI: 0,78; 2,51). Das würde bedeuten, dass die Behandlungen die Wahrscheinlichkeit, wieder arbeiten zu können, um 40% erhöhen könnten.

Bei der zusätzlich durchgeführten Analyse der Wahrscheinlichkeitsdifferenz ergab sich ein absoluter Vorteil für die aktive Behandlung von 12% (95%-KI: -0,04%; 29%). Das Ergebnis war jedoch mit p=0,15 statistisch nicht signifikant.

3.4.7.3 Vergleich der Ergebnisse aus den Abschnitten 3.4.7.1 und 3.4.7.2

Obwohl die Analysen der Wiedereingliederung von arbeitsunfähigen Patienten in das Arbeitsleben eine klare Aussage nicht zuließen, so war die relative Wiederherstellung der Arbeitsfähigkeit (40%) der Vermeidung von Arbeitsunfähigkeit mit 41% sehr ähnlich.
Dies zeigt, dass die Behandlungen sowohl die Arbeitsfähigkeit erhalten konnten (Abschnitt 3.4.7.1) als auch eventuell helfen könnten, Arbeitsunfähigkeit zu verringern (Abschnitt 3.4.7.1).

3.4.7.4 Meta-Analyse zu der Arbeitsfähigkeit zu Studienende

Abschließend wurde untersucht, ob sich die absolute Anzahl an arbeitsfähigen Patienten in den Behandlungsgruppen zum Ende hin unterschied. Aus den Arbeiten von Carlsson und Sjölund 2001, Manchikanti et al. 2010 und Smolen et al. 2006 konnten derartige Daten extrahiert werden (Abbildung 21).

	Active		Control			Risk Ratio
Study or Subgroup	**Events**	**Total**	**Events**	**Total**	**Weight**	**IV, Fixed, 95% CI**
Carlsson & Sjölund 2001	18	21	6	11	2.7%	1.57 [0.89, 2.77]
Manchikanti 2010	22	60	22	60	3.9%	1.00 [0.62, 1.60]
Smolen 2006	454	612	163	231	93.5%	1.05 [0.96, 1.16]
Total (95% CI)		**693**		**302**	**100.0%**	**1.06 [0.97, 1.16]**
Total events	494		191			

Heterogeneity: Chi² = 1.94, df = 2 (P = 0.38); I² = 0%
Test for overall effect: Z = 1.25 (P = 0.21)

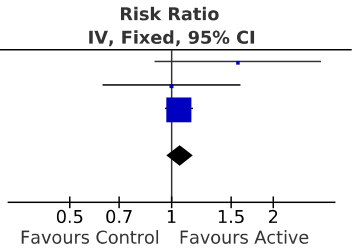


Abbildung 21: *Meta-Analyse der Wahrscheinlichkeitsquotienten zu Studienende arbeitsfähig zu sein.*

Für den RR fand sich ein nicht statistisch signifikanter Wert von 1,06 (95%-KI: 0,97; 1,16).

Der gepoolte RD aus einer zusätzlichen Analyse stellte sich mit einer durchschnittlichen absoluten Verbesserung von 4% (95%-KI: -2%; 10%) dar. Wieder konnte jedoch keine Signifikanz erreicht werden.

3.5 Schmerzbezogene Endpunkte

Von den in Tabelle 1 gelisteten 27 Patientenkollektiven führten nur Carlsson und Sjölund 2001 und Manchikanti et al. 2010 keinen Signifikanztest ihrer arbeitsbezogenen Ergebnisse durch. Von den übrigen 25 berichteten 16 über Schmerzendpunkte (Albert et al. 2013, Baron et al. 2010, Bennett et al. 2003, Bradley et al. 2010, Chappell et al. 2009, Chappell et al. 2011, Egsmose et al. 1997, Kavanaugh et al. 2006, Licciardone et al. 2003, Markenson et al. 2005, Meireles et al. 2010, Skljarevski et al. 2009, Skljarevski et al. 2010a, Skljarevski et al. 2010b, Straube et al. 2011, van der Heijde et al. 2006). Die übrigen neun erfüllten diese Bedingung nicht (Barkham et al. 2010, Bejarano et al. 2008, He et al. 2005, Jarzem et al. 2005, Kavanaugh et al. 2009a, Kavanaugh et al. 2013, Lehmann et al. 1986, Smolen et al. 2006, Strand et al. 1999). Um Doppelzählungen von Patienten zu vermeiden, wurden Arnold et al. 2005, Arnold et al. 2009, Bennett et al. 2005 und Chappell et al. 2008 hier nicht berücksichtigt. Die folgenden schmerzbezogenen Endpunkte wurden berichtet:

- Acht Studien erfassten den Unterpunkt „körperliche Schmerzen“ der „Short Form 36 Health Survey Questionnaire“
  (Bennett et al. 2003, Bradley et al. 2010, Chappell et al. 2009, Chappell et al. 2011, Licciardone et al. 2003, Skljarevski et al. 2009, Skljarevski et al. 2010a, Skljarevski et al. 2010b).
- Sieben erhoben eine visuelle Analogskala für Schmerzintensität
  (Bennett et al. 2003, Egsmose et al. 1997, Kavanaugh et al. 2006, Licciardone et al. 2003, Meireles et al. 2010, Straube et al. 2011, van der Heijde et al. 2006).
- Sechs untersuchten mindestens einen relevanten Unterpunkt des „Brief Pain Inventory“
  (Chappell et al. 2009, Chappell et al. 2011, Markenson et al. 2005, Skljarevski et al. 2009, Skljarevski et al. 2010a, Skljarevski et al. 2010b).

- Drei berichteten über Schmerzen im Rahmen des „Fibromyalgia Impact Questionnaire“
  (Bennett et al. 2003, Bradley et al. 2010, Straube et al. 2011).
- Albert et al. 2013 verwendeten eine kontinuierliche Skala von 0 bis 10 für Rückenschmerzen.
- Baron et al. 2010 erfassten eine allgemeine Schmerzskala von 0 bis 10.

3.5.1 Meta-Analysen der schmerzbezogenen Endpunkte

Insgesamt konnten aus 13 Studien geeignete Daten zur Erstellung von Meta-Analysen über schmerzbezogene Endpunkte extrahiert werden (Tabelle 6).

	schmerzbezogener Endpunkt
Albert 2013	Low Back Pain Rating Scale
Bennett 2005	FIQ-P
Bradley 2010	FIQ-P
Chappell 2009	BPI-P
Chappell 2011	BPI-P
Egsmose 1997	VAS
Kavanaugh 2006	VAS
Markenson 2005	BPI-P
Skljarevski 2009	BPI-P
Skljarevski 2010a	BPI-P
Skljarevski 2010b	BPI-P
Straube 2011	VAS
van der Heijde 2006	VAS

Tabelle 6: *Schmerzbezogene Endpunkte der Studien*
BPI-P= Brief Pain Inventory – Average Pain; FIQ-P = Fibromyalgia Impact Questionnaire – Pain;
VAS = Visual Analogue Scale for Pain

3.5.1.1 Meta-Analyse der Durchschnittsdifferenzen zu schmerzbezogenen Endpunkten

Zehn Studien berichteten in ausreichendem Maße über die Differenz zwischen Studienbeginn und -ende eines Schmerzendpunktes (Bradley et al. 2010, Chappell et al. 2009, Chappell et al. 2011, Egsmose et al. 1997, Kavanaugh et al. 2006, Markenson et al. 2005, Skljarevski et al. 2009, Skljarevski et al. 2010a, Skljarevski et al. 2010b, van der Heijde et al. 2006). Die Daten über schmerzbezogene Endpunkte für die Studien von Kavanaugh et al. 2006 und van der Heijde et al. 2006 entstammen den früheren Publikationen von Antoni et al. 2005 und van der Heijde et al. 2005. Die für die Analyse untersuchten Endpunkte für diese und die darauf folgende Meta-Analyse finden sich in Tabelle 6.

Die Meta-Analyse der standardisierten Behandlungseffekte wurde mit einem *random effects model* erstellt (Abbildung 22).

	Active			Control				Std. Mean Difference
Study or Subgroup	**Mean**	**SD**	**Total**	**Mean**	**SD**	**Total**	**Weight**	**IV, Random, 95% CI**
Bradley 2010	-1.96	2.49	796	-1.3	2.24	533	13.4%	-0.28 [-0.39, -0.17]
Chappell 2009	-2.82	2.17	107	-1.85	2.26	116	9.8%	-0.44 [-0.70, -0.17]
Chappell 2011	-2.51	2.26	128	-1.72	2.04	128	10.3%	-0.37 [-0.61, -0.12]
Egsmose 1997	-30.9	23.11	37	-29.2	22.68	42	6.3%	-0.07 [-0.52, 0.37]
Kavanaugh 2006	-39.6	56.1	100	11.8	109.8	100	9.4%	-0.59 [-0.87, -0.30]
Markenson 2005	-1.7	2.22	55	-0.6	2.86	51	7.3%	-0.43 [-0.81, -0.04]
Skljarevski 2009	-2.33	2.29	272	-1.87	2.34	113	10.9%	-0.20 [-0.42, 0.02]
Skljarevski 2010a	-2.08	2.11	111	-1.3	2.05	116	9.9%	-0.37 [-0.64, -0.11]
Skljarevski 2010b	-2.25	2.09	195	-1.65	2.12	199	11.4%	-0.28 [-0.48, -0.09]
van der Heijde 2006	-2.9	3.56	199	-0.3	1.93	199	11.3%	-0.91 [-1.11, -0.70]
Total (95% CI)			**2000**			**1597**	**100.0%**	**-0.40 [-0.55, -0.25]**

Heterogeneity: Tau² = 0.04; Chi² = 36.05, df = 9 (P < 0.0001); I² = 75%
Test for overall effect: Z = 5.34 (P < 0.00001)

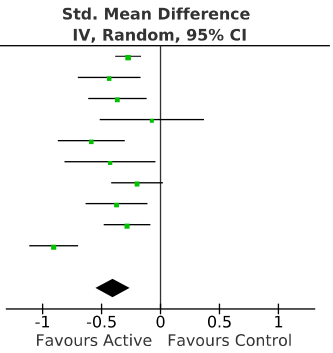


Abbildung 22: *Meta-Analyse von Schmerzendpunkten über alle Studien hinweg, die Ergebnisse als Durchschnittsdifferenz berichteten. Durchschnittlicher Behandlungseffekt ausgedrückt als standardisierte Mittelwertsdifferenz.*

Der standardisierte Unterschied zwischen den Behandlungsgruppen betrugt -0,40 (95%-KI: -0,55; -0,25).

3.5.1.2 Meta-Analyse der Enddaten zu schmerzbezogenen Endpunkten

Albert et al. 2013 und Bennett et al. 2003 blieben den Bericht der Standardabweichungen der einzelnen Gruppenunterschiede schuldig. So konnten nur ihre finalen Endpunktwerte und die der Studien von Chappell et al. 2009, Markenson et al. 2005 sowie von dem Erstautor der Meta-Analyse von Straube et al. 2011 bereitgestellten Endpunktdaten miteinander verglichen werden (Abbildung 23). Die Endpunkte der einzelnen Studien können in Tabelle 6 eingesehen werden.

	Active			Control				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Albert 2013	3.7	3.33	100	6.3	2.74	100	19.1%	-0.85 [-1.14, -0.56]
Bennett 2003	5.7	2.7	156	6.4	2.5	157	21.1%	-0.27 [-0.49, -0.05]
Chappell 2009	3.48	2.16	107	4.48	2.28	116	19.8%	-0.45 [-0.71, -0.18]
Markenson 2005	4.9	2.22	55	6	2.86	51	16.2%	-0.43 [-0.81, -0.04]
Straube 2011	52.89	26.06	1350	55.7	25.51	466	23.9%	-0.11 [-0.21, -0.00]
Total (95% CI)			**1768**			**890**	**100.0%**	**-0.40 [-0.66, -0.14]**

Heterogeneity: Tau² = 0.07; Chi² = 26.34, df = 4 (P < 0.0001); I² = 85%
Test for overall effect: Z = 3.01 (P = 0.003)

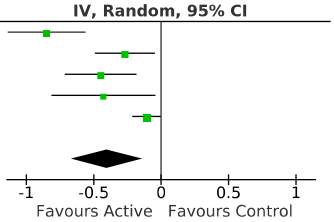


Abbildung 23: *Meta-Analyse von Schmerzendpunkten über alle Studien hinweg, die Daten zum Studienende berichteten. Durchschnittlicher Behandlungseffekt ausgedrückt als standardisierte Mittelwertsdifferenz.*

Das Ergebnis lag mit einem standardisierten Behandlungseffekt von -0,40 (95%-KI: -0,66; -0,14) nah an der vorangegangen Analyse. Es lag eine relativ hohe Heterogenität von I^2 = 85% vor.

3.5.1.3 Heterogenitäts- und Sensitivitätsanalyse

Der aus den Daten der Meta-Analyse aus Abschnitt 3.5.1.1 erstellte Funnel-Plot (Abbildung 24) zeigte, dass die Ergebnisse der Studien symmetrisch um die gepoolte Mittelwertsdifferenz angeordnet sind. Nur die Ergebnisse von van der Heijde et al. 2006 wichen stark von den Ergebnissen der anderen Studien ab. Der Heterogenitätswert von 75% ist ebenfalls auf die Studie von van der Heijde et al. 2006 zurück zu führen. Der Grund hierfür könnte in der Extraktion der Daten

dieser Studie liegen. Dies wird im Detail in Abschnitt 4.1 diskutiert. Eine Meta-Analyse ohne die Daten von van der Heijde et al. 2006 verringerte den gepoolten Behandlungseffekt auf -0,31 (95%-KI: -0,39; -0,24). Die Heterogenität zwischen den Studien sank so auf I^2= 0%.

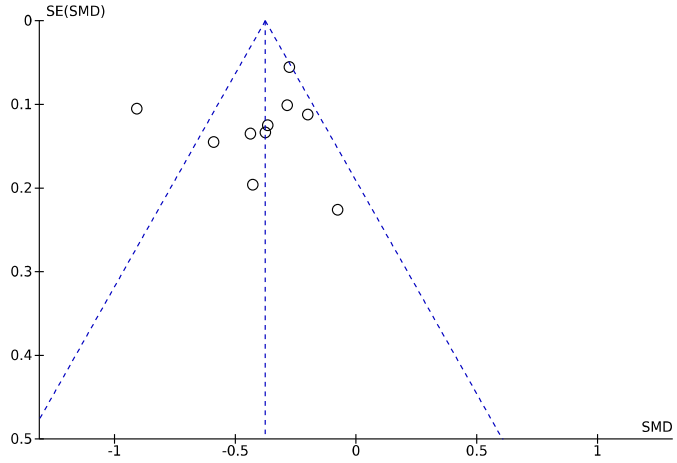


Abbildung 24: *Funnel-Plot der Studien, welche sich mit Schmerzendpunkten beschäftigen*

3.5.2 Vergleich arbeitsbezogener Endpunkte mit Schmerzendpunkten

Werden nun die Behandlungseffekte der schmerzbezogenen Endpunkte aus Abschnitt 3.5.1.1 (standardisierter Effekt -0,40) mit den Ergebnissen der Meta-Analyse der arbeitsbezogenen Endpunkte aus Abschnitt 3.4.1.1 (standardisierter Effekt -0,35) verglichen, so ist bereits erkennbar, dass beide Endpunkte ähnlich stark von den Behandlungen profitierten. Um den Zusammenhang zwischen arbeits- und schmerzbezogenen Endpunkten genauer zu untersuchen, wurden in den nachfolgenden Abschnitten Regressionen erstellt.

3.5.2.1 Regression zu schmerz- und arbeitsbezogenen Endpunkten des „Brief Pain Inventory“

Sieben Studien berichteten mithilfe des „Brief Pain Inventory“ in sehr ähnlicher Weise über schmerzbezogene und arbeitsbezogene Endpunkte (Arnold et al. 2005, Chappell et al. 2009, Skljarevski et al. 2009, Chappell et al. 2011, Skljarevski et al. 2010a , Skljarevski et al. 2010b, Markenson et al. 2005). Zum direkten Vergleich wurden nun die Mittelwertsdifferenzen zwischen den Behandlungsgruppen der Unterpunkte „Durchschnittsschmerz“ (*average pain*) und „Beeinträchtigung bei der Arbeit“ (*interference with normal work*) untersucht. Beide Teilfragen über selbst empfundene Schmerzen, beziehungsweise selbst empfundene Beeinträchtigung bei der Arbeit, werden bei der BPI mit einer Skala erfasst, welche 0 als keine Beschwerden beschreibt und 10 als am schlimmsten vorstellbare Schmerzen, beziehungsweise Beeinträchtigung, charakterisieren. Die Ergebnisse beider Teilfragen wurden standardisiert und mithilfe von in Abschnitt 2.4.7 beschriebenen linearen Modellen untersucht (Abbildung 25).

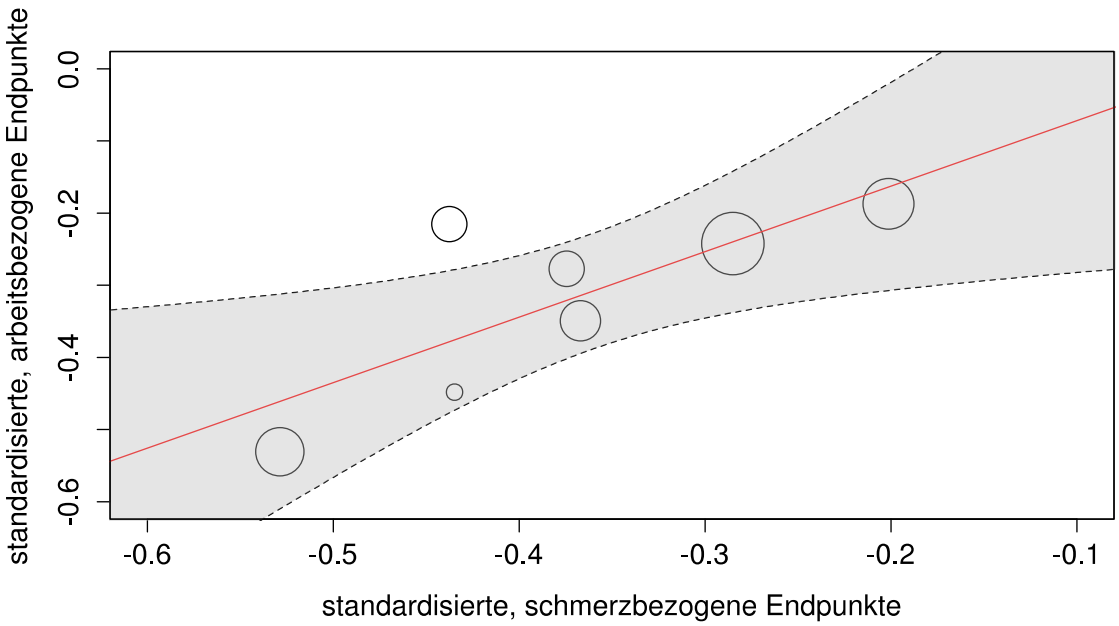


Abbildung 25: *Regression von standardisierten, schmerz- und arbeitsbezogenen Endpunkten des „Brief Pain Inventory“. Die Größe der Punkte repräsentiert die individuelle Gewichtung der einzelnen Studien basierend auf ihrer inversen Varianz.*
Die gestrichelten Linien markieren das 95%-Konfidenzintervall.

Die Regressionsanalyse zeigte einen statistisch signifikanten (p=0,03), positiven linearen Zusammenhang der standardisierten Intergruppenunterschiede. Die mit 0,91 an 1 grenzende Steigung der Regressionsgeraden (rote Gerade) lässt darauf schließen, dass Schmerzendpunkte und arbeitsbezogene Endpunkte fast gleich auf Behandlung reagieren, wenn die Messinstrumente wie in diesem Fall (BPI) ähnlich gestaltet sind.

3.5.2.2 Regression zu allen schmerz- und arbeitsbezogenen Endpunkten

Drei weitere Arbeiten berichteten die Mittelwertsdifferenzen ihrer arbeits- und schmerzbezogenen Endpunkte und wurden mit den Studien aus Abschnitt 3.5.2.1 zusammen analysiert. Bradley et al. 2010 untersuchte diese mithilfe von Teilfragen des „Fibromyalgia Impact Questionnaire“. Kavanaugh et al. 2006 und van der Heijde 2006 erhoben die Einschränkungen bei der Produktivität und die empfundene Schmerzintensität mit visuellen Analogskalen. Die Daten bezüglich der schmerzrelevanten Endpunkte für Kavanaugh et al. 2006 und van der Heijde et al. 2006 entstammen wieder den Publikationen von Antoni et al. 2005 und van der Heijde et al. 2005. Um überlappende Patientenpopulationen durch den Einschluss der Arbeit von Bradley et al. 2010 zu vermeiden, wurde die Arbeit von Arnold et al. 2005 ausgeschlossen. Die Ergebnisse wurden standardisiert und in einer Regression abgetragen (Abbildung 26).

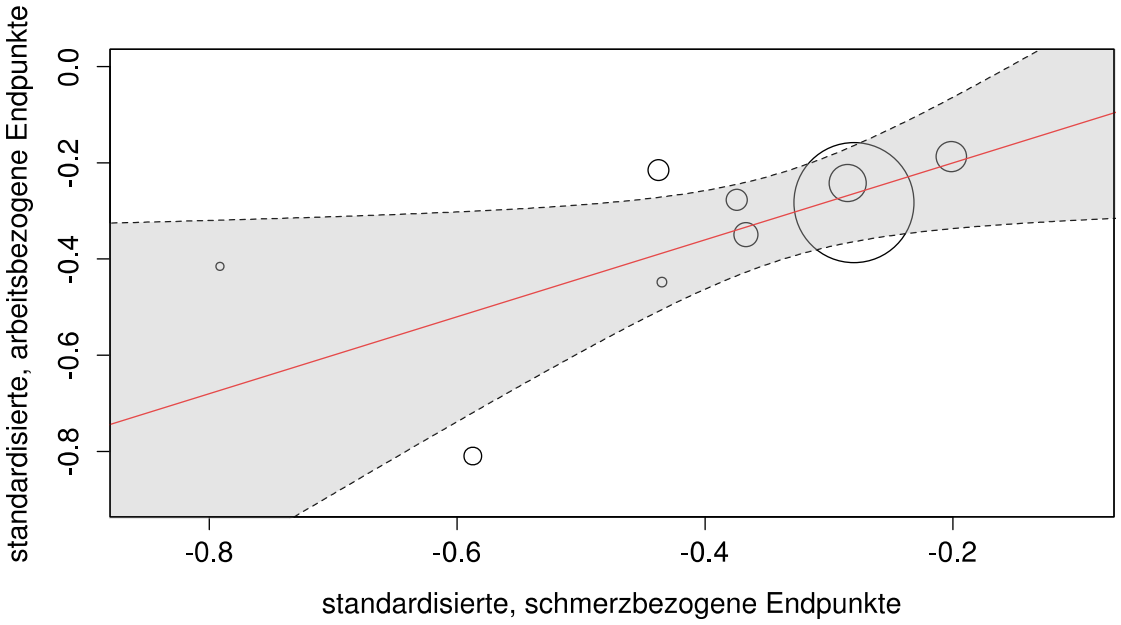


Abbildung 26: *Regression von standardisierten, schmerz- und arbeitsbezogenen Endpunkten. Die Größe der Punkte repräsentiert die individuelle Gewichtung der einzelnen Studien basierend auf ihrer inversen Varianz.*
Die gestrichelten Linien markieren das 95%-Konfidenzintervall.

Die Steigung der statistisch signifikanten Regressionsgeraden betrug 0,80 (p = 0.04). Wie in Abschnitt 3.5.2.1 konnte ein statistisch signifikanter Zusammenhang zwischen der Verbesserung von arbeitsbezogenen Endpunkten und der Verbesserung von Schmerzendpunkten nachgewiesen werden.

3.6 Ansprechen auf Behandlungen

In den letzten Jahren hat sich die Angabe des Anteils der Patienten, die eine Verbesserung ihres Schmerzempfindens erfahren, als Standardmessinstrument des Behandlungserfolges bei chronischen Schmerzerkrankungen etabliert (Moore et al. 2010a). Dabei werden die kontinuierlichen Werte der Endpunkte in ein binäres Bewertungssystem überführt. Hierzu werden die Anteile, das heißt die Ansprechraten der Patienten mit einem Behandlungserfolg von mindestens 30% sowie 50% im Bereich Schmerzverbesserung, erfasst. Weitere Ausführungen in Bezug auf Hintergründe und zu Grunde liegende Verteilungen sind in Abschnitt 4.5 diskutiert. Ebenso werden im Bereich der Rheumatologie häufig Ansprechraten in Bezug auf die Kriterien des American College of Rheumatology (ACR) erhoben. Diese definieren für jede rheumatische Erkrankung spezifische Befunde, um ein möglichst homogenes Patientenkontingent über alle Studien hinweg zu gewährleisten. Beispiele hierfür sind in Abschnitt 1.4.2 zu Fibromyalgie oder in Abschnitt 1.4.3 zu der rheumatoiden Arthritis gegeben.

3.6.1 Meta-Analysen der Ansprechraten für schmerzbezogene Endpunkte

Über schmerzbezogene Ansprechraten berichteten insgesamt neun Publikationen:

- Fünf Studien berichteten über den Unterpunkt „Durchschnittsschmerz“ des „Brief Pain Inventory“
  (Arnold et al. 2009, Chappell et al. 2011, Markenson et al. 2005, Skljarevski et al. 2010a, Skljarevski et al. 2010b).
- Zwei verwendeten eine Linkert-Skala (0 bis 10 Punkte) für 24-stündigen Durchschnittsschmerz, angelehnt an das BPI
  (Chappell et al. 2009, Skljarevski et al. 2009).

- Bennett et al. 2003 untersuchten eine visuelle Analogskala für Durchschnittsschmerzen.
- Für die in der Meta-Analyse von Straube et al. 2011 untersuchten Studien fand sich eine weitere Veröffentlichung, die über dieselben Studien sowie über die Ansprechraten des Unterpunktes „körperliche Schmerzen“ des SF-36 berichtete (Straube et al. 2010).

Die Arbeit von Arnold et al. 2009 entstammte aus derselben Arbeitsgruppe wie die Meta-Analyse von Bradley et al. 2010 und berichtete über dieselben Studien. Sie wurde hier wegen ihrer verständlicheren Berichterstattung der Ansprechraten anstatt der Arbeit von Bradley et al. 2010 verwendet.
Bis auf Skljarevski et al. 2009 berichteten alle Studien über Ansprechraten sowohl für 30%ige (30%-Ansprechraten) als auch 50%ige (50%-Ansprechraten) Schmerzverbesserungen. Nur Skljarevski et al. 2009 berichteten nicht über die 50%-Ansprechraten. Insgesamt wurden 5849 Patienten untersucht. Die 30%-Ansprechraten der Kontrollgruppen variierten zwischen 23,6% und 48,7%, die der 50%-Ansprechraten waren zwischen 18,5% und 34,7% angesiedelt.
Die Ansprechraten wurden, wie in Abschnitt 2.4.6 beschrieben, mithilfe von erstellten Wahrscheinlichkeitsquotienten und Wahrscheinlichkeitsdifferenzen in den nachfolgenden Meta-Analysen untersucht.

3.6.1.1 Meta-Analysen der 30%-Ansprechraten

Die Meta-Analyse der Wahrscheinlichkeitsquotienten der 30%-Ansprechraten erfolgte unter Anwendung des Random-Effekts-Model (Abbildung 27).

	Active		Control			Risk Ratio
Study or Subgroup	**Events**	**Total**	**Events**	**Total**	**Weight**	**IV, Random, 95% CI**
Arnold 2009	369	774	169	526	14.5%	1.48 [1.28, 1.71]
Bennett 2003	65	156	37	157	8.7%	1.77 [1.26, 2.48]
Chappell 2009	63	107	52	116	11.0%	1.31 [1.02, 1.70]
Chappell 2011	84	128	56	128	11.7%	1.50 [1.19, 1.89]
Markenson 2005	21	56	9	51	3.5%	2.13 [1.07, 4.21]
Skljarevski 2009	151	287	51	117	11.7%	1.21 [0.96, 1.52]
Skljarevski 2010a	59	111	46	116	10.2%	1.34 [1.01, 1.78]
Skljarevski 2010b	111	195	97	199	13.2%	1.17 [0.97, 1.41]
Straube 2011	759	1936	269	689	15.5%	1.00 [0.90, 1.12]
Total (95% CI)		**3750**		**2099**	**100.0%**	**1.33 [1.15, 1.53]**
Total events	1682		786			

Heterogeneity: Tau² = 0.03; Chi² = 30.51, df = 8 (P = 0.0002); I² = 74%
Test for overall effect: Z = 3.89 (P < 0.0001)

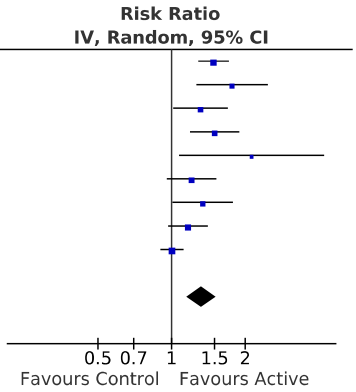


Abbildung 27: *Meta-Analyse der Wahrscheinlichkeitsquotienten der 30%-Schmerzansprechraten.Die Daten für das Patientenkollektiv von Straube et al. 2011 entstammen der Publikation von Straube et al. 2010.*

Aus der Analyse geht hervor, dass es 33% (95%-KI: 15%; 53%) wahrscheinlicher war, eine mindestens 30%ige Schmerzminderung zu erfahren, als in der Kontrollgruppe.

Da die relativen Wahrscheinlichkeitsquotienten nichts über den absoluten Behandlungserfolg aussagen, wurde weiterhin eine Meta-Analyse der Wahrscheinlichkeitsdifferenzen der 30%-Ansprechraten erstellt. Die gepoolte Wahrscheinlichkeitsdifferenz zwischen aktiver Gruppe und Kontrollgruppe waren mit 13% (95%-KI: 6%; 19%) statistisch signifikant.

3.6.1.2 Meta-Analysen der 50%-Ansprechraten

Wie bereits in Abschnitt 3.6.1.1 wurde auch bei der Meta-Analyse der 50%-Ansprechraten das *random effects model* verwendet (Abbildung 28).

Study or Subgroup	Active Events	Active Total	Control Events	Control Total	Weight	Risk Ratio IV, Random, 95% CI
Arnold 2009	273	774	117	526	17.4%	1.59 [1.32, 1.91]
Bennett 2003	55	156	28	157	10.4%	1.98 [1.33, 2.94]
Chappell 2009	51	107	34	116	11.9%	1.63 [1.15, 2.30]
Chappell 2011	56	128	41	128	12.7%	1.37 [0.99, 1.88]
Markenson 2005	11	56	3	51	1.9%	3.34 [0.99, 11.30]
Skljarevski 2010a	43	111	31	116	10.8%	1.45 [0.99, 2.12]
Skljarevski 2010b	95	195	69	199	15.5%	1.41 [1.11, 1.78]
Straube 2011	643	1936	214	689	19.3%	1.07 [0.94, 1.22]
Total (95% CI)		**3463**		**1982**	**100.0%**	**1.46 [1.22, 1.74]**
Total events	1227		537			

Heterogeneity: Tau² = 0.04; Chi² = 22.27, df = 7 (P = 0.002); I² = 69%
Test for overall effect: Z = 4.15 (P < 0.0001)

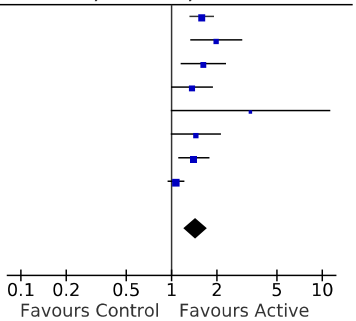


Abbildung 28: *Meta-Analyse der Wahrscheinlichkeitsquotienten der 50%-Schmerzansprechraten. Die Daten für das Patientenkollektiv von Straube et al. 2011 entstammen der Publikation von Straube et al. 2010.*

Die gepoolte Wahrscheinlichkeit, eine 50%ige Verbesserung zu erreichen, war in den Behandlungsgruppen 46% (95%-KI: 22%; 74%) größer als in den Kontrollgruppen.

In einer weiteren Meta-Analyse betrug die Wahrscheinlichkeitsdifferenz zwischen den beiden Gruppen 12% (95%-KI: 7%; 17%). Beide Analysen konnten mit einen p-Wert von unter 0,00001 ein statistisch signifikantes Ergebnis vorweisen.

3.6.1.3 Heterogenitäts- und Sensitivitätsanalysen

Die Meta-Analyse der Wahrscheinlichkeitsquotienten der 30%-Ansprechraten in Abschnitt 3.6.1.1 wies einen recht hohen Heterogenitätsbefund (I^2 = 74%) auf (Abbildung 27). Werden die Eigenschaften und Themenbereiche der einzelnen Studien betrachtet, so fällt auf, dass die Arbeiten von Bennett et al. 2003, Markenson et al. 2005 und Straube et al. 2011 die einzigen sind, die nicht mit Duloxetin behandelten, und daher möglicherweise andere Effekte gemessen wurden als in den restlichen Studien (Tabelle 7).

	Erkrankung	**Behandlung**
Arnold 2009	Fibromyalgie	Duloxetin
Bennett 2003	Fibromyalgie	Paracetamol-Tramadol
Chappell 2009	Osteoarthritis/Knie	Duloxetin
Chappell 2011	Osteoarthritis/Knie	Duloxetin
Markenson 2005	Osteoarthritis	Oxicodon
Skljarevski 2009	chr. Rückenschmerzen	Duloxetin
Skljarevski 2010a	chr. Rückenschmerzen	Duloxetin
Skljarevski 2010b	chr. Rückenschmerzen	Duloxetin
Straube 2011	Fibromyalgie	Pregabalin

Tabelle 7: *Die Studien der Meta-Analyse mit ihren untersuchten Erkrankungen und ihren verwendeten Behandlungen.*

Eine zusätzlich erstellte Sensitivitätsanalyse, bei der nur die Duloxetin-Studien untersucht wurden, zeigte einen durchschnittlichen RR von 1,34 (95%-KI: 1,23; 1,53) mit einer geringen Heterogenität von I^2 = 12%. Die Analyse der RD ergab 14% (95%-KI: 10%; 10%) mit einem Heterogenitätswert von I^2 = 0%.

Ebenso ließ sich bei der Meta-Analyse des RR der 50%-Ansprechraten in Abschnitt 3.6.1.2 mit I^2= 69% eine hohe Heterogenität feststellen. Wie auch zuvor sind hier andere Behandlungseffekte bei Bennett et al. 2003, Markenson et al. 2005 und Straube et al. 2011 zu vermuten. Bei einer Analyse der Duloxetin-Studien ergab sich ein durchschnittlicher RR von 1,50 (95%-KI: 1,33; 1,69) und ein absoluter RD von 0,13 (95%-KI: 0,10; 0,17). Es konnte keine Heterogenität festgestellt werden.

3.6.2 Meta-Analysen der ACR-Ansprechraten

Im Bereich der Rheumatologie werden häufig Ansprechraten in Bezug auf die Kriterien des American College of Rheumatology (ACR) erhoben. Diese definieren für jede rheumatische Erkrankung spezifische Befunde, um ein möglichst homogenes Patientenkontingent über alle Studien hinweg zu gewährleisten.

Daten zu ACR-Ansprechraten konnten in 6 Studien gefunden werden. Drei berichteten direkt über diesen Endpunkt (Bejarano et al. 2008, Kavanaugh et al. 2009a, Strand et al. 1999). Drei weitere Studien berichteten selbst nicht darüber. Sie basierten jedoch auf früher veröffentlichten Arbeiten, die diesen Endpunkt untersucht hatten. So konnten Informationen für Kavanaugh et al. 2006 in Antoni et al. 2005, für Smolen et al. 2006 in St. Clair et al. 2004 und für Kavanaugh et al. 2013 in Kavanaugh et al. 2009b gefunden werden.

Zumeist wurde jeweils die Anteile der Patienten erfasst, die nach Studienende eine 20%ige, 50%ige oder 70%ige Verbesserung nach den Kriterien des ACR vorzuweisen hatten. Nur Strand et al. 1999 berichteten nicht über 70%-Ansprechraten. Insgesamt wurden in den ACR-Studien 3135 Patienten erfasst.

In Tabelle 8 sind die Ansprechraten der Kontrollgruppen dokumentiert.

	ACR-20%	ACR-50%	ACR-70%
Ansprechen der Kontrollgruppen	11,0% - 54,8%	3,0% - 45,2%	1,0% - 37,5%

Tabelle 8: *Spannweite der Ansprechraten der Kontrollgruppen*

Wie bereits in Abschnitt 3.6.1 wurden Meta-Analysen der Studienergebnisse, ausgedrückt als RR und RD, erstellt.

3.6.2.1 Meta-Analysen der 20%-ACR-Ansprechraten

Die berichteten 20%-Ansprechraten wurden als RR ausgedrückt und mithilfe eines *random effects models* gepoolt (Abbildung 29).

Study or Subgroup	Active Events	Active Total	Control Events	Control Total	Weight	Risk Ratio IV, Random, 95% CI
Bejarano 2008	54	75	40	73	18.1%	1.31 [1.02, 1.69]
Kavanaugh 2006	58	100	11	100	14.1%	5.27 [2.95, 9.44]
Kavanaugh 2009a	468	783	27	199	17.0%	4.41 [3.09, 6.28]
Kavanaugh 2013	165	292	14	113	15.2%	4.56 [2.76, 7.52]
Smolen 2006	454	706	71	274	18.5%	2.48 [2.02, 3.05]
Strand 1999	156	319	26	101	17.1%	1.90 [1.34, 2.70]
Total (95% CI)		**2275**		**860**	**100.0%**	**2.84 [1.86, 4.34]**
Total events	1355		189			

Heterogeneity: Tau² = 0.24; Chi² = 48.34, df = 5 (P < 0.00001); I² = 90%
Test for overall effect: Z = 4.84 (P < 0.00001)

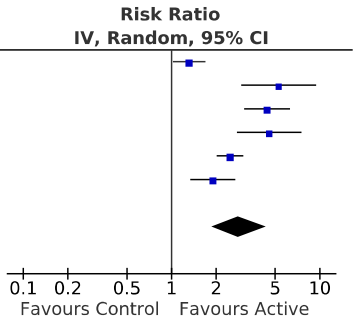


Abbildung 29: *Meta-Analyse der Wahrscheinlichkeitsquotienten der 20%-ACR-Ansprechraten. Die Daten für die Patientenkollektive von Kavanaugh et al. 2006, Smolen et al. 2006, Kavanaugh et al. 2013 entstammen den Publikationen von Antoni et al. 2005, St.Clair et al. 2004 und Kavanaugh et al. 2009b*

Die Analyse zeigte, dass es in der aktiven Behandlungsgruppe 184% wahrscheinlicher war, eine 20%-Ansprechrate zu erreichen, als in der Kontrollgruppe (95%-KI: 86%; 334%).

In einer weiteren Analyse der Wahrscheinlichkeitsdifferenzen erreichten durchschnittlich 37% (95%-KI: 29%; 45%) mehr Patienten der aktiven Behandlungsgruppen eine 20%-Verbesserung als die Kontrollgruppe.

3.6.2.2 Meta-Analysen der 50%-ACR-Ansprechraten

Die 50%-Ansprechraten wurden ebenfalls mit einem *random effects model* untersucht (Abbildung 30).

Study or Subgroup	Active Events	Active Total	Control Events	Control Total	Weight	Risk Ratio IV, Random, 95% CI
Bejarano 2008	42	75	33	73	19.4%	1.24 [0.90, 1.71]
Kavanaugh 2006	36	100	3	100	11.8%	12.00 [3.82, 37.70]
Kavanaugh 2009a	301	783	15	199	18.1%	5.10 [3.11, 8.36]
Kavanaugh 2013	102	292	5	113	14.4%	7.89 [3.30, 18.86]
Smolen 2006	339	706	88	274	20.2%	1.50 [1.24, 1.81]
Strand 1999	91	319	8	101	16.2%	3.60 [1.81, 7.16]
Total (95% CI)		**2275**		**860**	**100.0%**	**3.37 [1.84, 6.15]**
Total events	911		152			

Heterogeneity: Tau² = 0.46; Chi² = 51.43, df = 5 (P < 0.00001); I² = 90%
Test for overall effect: Z = 3.95 (P < 0.0001)

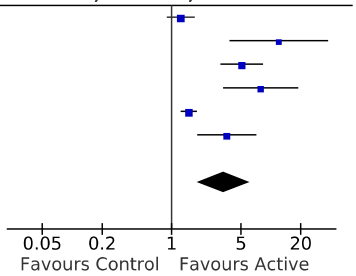


Abbildung 30: *Meta-Analyse der Wahrscheinlichkeitsquotienten der 50%-ACR-Ansprechraten. Die Daten für die Patientenkollektive von Kavanaugh et al. 2006, Smolen et al. 2006, Kavanaugh et al. 2013 entstammen den Publikationen von Antoni et al. 2005, St.Clair et al. 2004 und Kavanaugh et al. 2009b*

Für die Analyse des RR zeigte sich, dass es für einen Patienten in den Behandlungsgruppen 237% wahrscheinlicher war, eine 50%-Verbesserung zu erreichen, als für Patienten in den Kontrollgruppen (95%-KI: 84%; 515%).

Die zusätzliche Meta-Analyse der Wahrscheinlichkeitsdifferenzen ergab einen gepoolten Unterschied von 25% (95%-KI: 18%; 31%).

3.6.2.3 Meta-Analysen der 70%-ACR-Ansprechraten

Auch die RR der 70%-Ansprechraten wurde mit einem *random effects model* analysiert (Abbildung 31).

Study or Subgroup	Active Events	Active Total	Control Events	Control Total	Weight	Risk Ratio IV, Random, 95% CI
Bejarano 2008	38	75	27	73	29.1%	1.37 [0.94, 1.99]
Kavanaugh 2006	15	100	1	100	8.8%	15.00 [2.02, 111.41]
Kavanaugh 2009a	164	783	6	199	22.5%	6.95 [3.12, 15.46]
Kavanaugh 2013	58	292	1	113	9.1%	22.45 [3.15, 160.12]
Smolen 2006	246	706	58	274	30.5%	1.65 [1.28, 2.11]
Total (95% CI)		**1956**		**759**	**100.0%**	**3.32 [1.65, 6.71]**
Total events	521		93			

Heterogeneity: Tau² = 0.40; Chi² = 24.13, df = 4 (P < 0.0001); I² = 83%
Test for overall effect: Z = 3.36 (P = 0.0008)

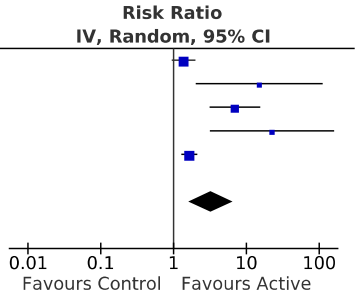


Abbildung 31: *Meta-Analyse der Wahrscheinlichkeitsquotienten der 70%-ACR-Ansprechraten. Die Daten für die Patientenkollektive von Kavanaugh et al. 2006, Smolen et al. 2006, Kavanaugh et al. 2013 entstammen den Publikationen von Antoni et al. 2005, St.Clair et al. 2004 und Kavanaugh et al. 2009b*

Die Meta-Analyse der 70%-Ansprechraten ergab gegenüber der Kontrollgruppe eine 232% (95%-KI: 65%; 571%) höhere Wahrscheinlichkeit für einen Erfolg in der aktiven Gruppe.

In einer weiteren Meta-Analyse der Wahrscheinlichkeitsdifferenzen lag der gepoolte Behandlungseffekt der Wahrscheinlichkeitsdifferenzen bei 17% (95%-KI: 14%; 19%).

3.6.2.4 Heterogenitäts- und Sensitivitätsanalysen

Die Ergebnisse der Studien bezüglich der ACR-Ansprechraten wiesen deutliche Unterschiede auf. So betrug der Heterogenitätswert I^2 der Analysen aus den Abschnitten 3.6.2.1, 3.6.2.2 und 3.6.2.3 oft über 80%. Dabei fällt auf, dass die drei Studien von Kavanaugh et al. aus 2006, 2009 und 2013 einen deutlich größeren relativen Behandlungseffekt aufwiesen, als die übrigen Studien. Im Funnel-Plot der 20%-ACR-Daten sind sie als die drei Werte rechts außerhalb des

Konfidenzintervalles zu erkennen (Abbildung 32). Es ist möglich, dass dies Anzeichen einer Veröffentlichungsverzerrung seien könnten, da Studien mit großem Effekt und großem Standardfehler berichtet wurden, jedoch keine Studien mit vergleichbarem Standardfehler aber einem niedrigeren Effekt erfasst werden konnten.

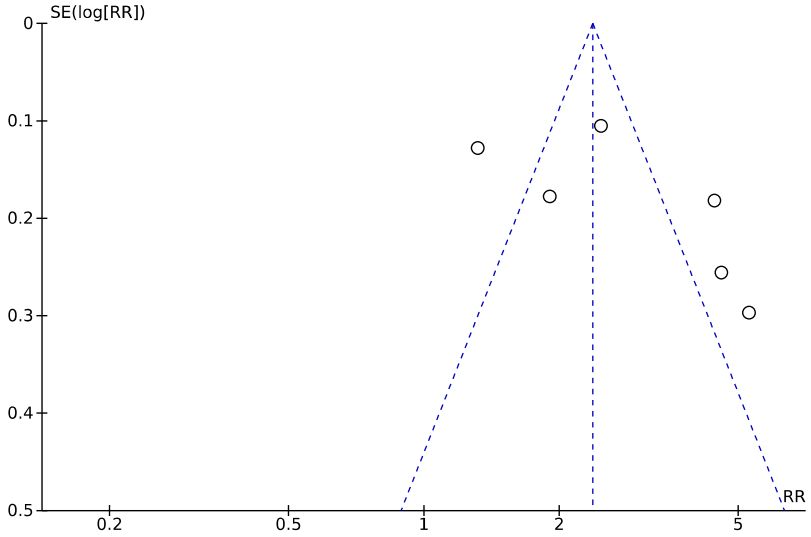


Abbildung 32: *Funnel-Plot der ACR-20% Meta-Analyse der Wahrscheinlichkeitsquotienten (RR)*

Vergleicht man die Ergebnisse der verwendeten Studien aus den Abschnitten 3.6.2.1, 3.6.2.2 und 3.6.2.3 mit den untersuchten Erkrankungen und Behandlungen (Tabelle 9), so lässt sich kein Zusammenhang zwischen den Studieneigenschaften und den Ergebnissen erkennen: Kavanaugh et al. 2006 und Kavanaugh et al. 2013 untersuchten als einzige Psoriasisarthritis (PA), jedoch übertreffen die Ergebnisse der rheumatoiden Arthrithis sogar die der PA (Kavanaugh et al. 2009a). Es ist daher eher ein Einfluss der Arbeitsgruppe, als ein Einfluss der Erkrankungen oder Behandlungen zu vermuten.

	Erkrankung	Behandlung
Bejarano 2008	rheumatoide Arthritis	Adalimumab & Prednisolon & Folsäure
Kavanaugh 2006	Psoriasisarthritis	Infliximab
Kavanaugh 2009	rheumatoide Arthritis	Certolizumab & MTX
Kavanaugh 2013	Psoriasisarthritis	Golimumab
Smolen 2006	rheumatoide Arthritis	Infliximab & MTX
Strand 1999	rheumatoide Arthritis	Leflunomid, Methotrexat

Tabelle 9: *Die Studien der Meta-Analyse mit ihren untersuchten Erkrankungen und ihren verwendeten Behandlungen.*

Weiterhin fiel auf, dass Tau^2 große Werte für die relativen Wahrscheinlichkeiten (RR) (0,24; 0,46; 0,40) annahm, während es für die absolute Differenz (RD) verschwindend gering war (0,01; 0,00; 0,00).
Die Varianz zwischen den unterschiedlichen Behandlungseffekten schien also weniger groß zu sein, wenn man statt der relativen Quotienten die absoluten Differenzen betrachtete. Wie in Abschnitt 4.2.3 beschrieben, würde man davon ausgehen, dass eher Verhältnisse statt absoluter Differenzen konstant blieben. Die Heterogenitätsbefunde und die Größe der Varianz Tau^2 sind in Tabelle 10 zusammengefasst.

	I^2 des RR	I^2 der RD	Tau^2 des RR	Tau^2 der RD
ACR-20%	90	81	0,24	0,01
ACR-50%	90	77	0,46	0,00
ACR-70%	83	0	0,40	0,00

Tabelle 10: *Gemessene Heterogenität (I^2-Wert in Prozent) und Varianz der unterschiedlichen Behandlungseffekte (Tau^2)*

Um die Zusammenhänge zwischen den Ansprechraten der Kontrollgruppen beziehungsweise der ermittelten Studienqualität mit den relativen und absoluten Ergebnissen der Analysen aus den Abschnitten 3.6.2.1, 3.6.2.2 und 3.6.2.3 genauer zu untersuchen, wurden lineare Regressionen durchgeführt. Tabelle 11 zeigt die Ergebnisse bezüglich der Ansprechraten der Kontrollgruppen.
Der Einfluss der Studienqualität wird in Tabelle 12 dargestellt.

Ansprechrate der Kontrollgruppe vs RD + RR

	Steigung	p	sig.
RD-20%	-0,007	0,03	*
RR-20%	-0,068	0,032	*
RD-50%	-0,005	0,023	*
RR-50%	-0,132	0,044	*
RD-70%	-0,002	0,207	
RR-70%	-0,176	0,278	

Tabelle 11: *Ergebnisse der linearen Regressionen der Ansprechraten der Kontrollgruppen mit den relativen und absoluten Studienergebnissen (RR/RD).*
*Statistisch signifikante Ergebnisse ($p < 0,05$) sind mit * gekennzeichnet.*

OQS vs RD + RR

	Steigung	p	sig.
RD-20%	-0,058	0,242	
RR-20%	-1,237	0,066	
RD-50%	-0,077	0,025	*
RR-50%	-2,561	0,036	*
RD-70%	-0,011	0,487	
RR-70%	-3,805	0,168	

Tabelle 12: *Ergebnisse der linearen Regressionen der Studienqualität mit den relativen und absoluten Studienergebnissen (RR/RD).*
*Statistisch signifikante Ergebnisse ($p < 0,05$) sind mit * gekennzeichnet.*

Es ließen sich keine für die Berichterstattungsart spezifischen Abhängigkeiten nachweisen. Konnte in einer Response-Kategorie (20%-, 50%-, 70%-ACR) eine Interaktion der Studienergebnisse mit der Reaktion der Kontrollgruppe oder der

Studienqualität gefunden werden, so waren immer beide Möglichkeiten der Präsentation, absolut (RD) und relativ (RR), für Verzerrungen anfällig.

Eine Abschätzung, welche Präsentationsform nun wegen erhöhter Robustheit gegenüber Verzerrungen vorzuziehen sei, ist anhand dieser Datenlage nicht möglich.

3.7 Ansprechraten und arbeitsbezogene Endpunkte

Bislang ist es für arbeitsspezifische Endpunkte nicht üblich, mithilfe von Ansprechraten Studienergebnisse zu untersuchen. Da bereits in den Analysen aus den Abschnitten 3.5.2.1 und 3.5.2.2 gezeigt werden konnte, dass schmerzbezogene und arbeitsbezogene Endpunkte unter den gleichen Voraussetzungen sehr ähnlich reagieren, wäre es naheliegend anzunehmen, dass dies bei Ansprechraten ebenfalls zutreffen könnte.
Daher wurde ein möglicher Zusammenhang zwischen den standardisierten Behandlungseffekten der arbeitsbezogenen Endpunkte sowie den relativen (RR) und absoluten (RD) Ergebnissen der kategorischen Schmerzverbesserungsraten untersucht.
Acht Studien beschrieben in adäquater Weise beide Endpunkte:

- Die Meta-Analyse von Arnold et al. 2009 berichtete über Ansprechraten bei Schmerzendpunkten (BPI: „Durchschnittsschmerz“) sowie über den Unterpunkt „Beeinträchtigung bei der Arbeit und in der Schule“ der SDS.
- Straube et al. verwendeten 2010 zur Berechnung der Ansprechraten für Schmerzendpunkte die Ergebnisse der Teilfrage des SF-36 „Bodily Pain“ und als arbeitsbezogenen Endpunkt untersuchten sie in einer Meta-Analyse im Jahr 2011 über dieselben Studien die Teilfrage „Beeinträchtigung bei der Arbeit“ des FIQ.
- Zwei Studien erfassten eine Linkert-Skala (0 bis 10 Punkte), angelehnt an das BPI für 24-stündigen Durchschnittsschmerz und als arbeitsbezogener Endpunkt wurde die Teilfrage „Beeinträchtigung bei der Arbeit“ des BPI verwendet (Chappell et al. 2009, Skljarevski et al. 2009).
- Die restlichen 4 Studien berichteten über die Teilfragen des BPI: „Beeinträchtigung bei der Arbeit“ und „Durchschnittsschmerz“ (Chappell et al. 2011, Markenson et al. 2005, Skljarevski et al. 2010a, Skljarevski et al. 2010b).

Der Zusammenhang der schmerzbezogenen Ansprechraten mit den arbeitsbezogenen Endpunkten werden in den folgenden Abschnitten untersucht. Regressionen der ACR-Ansprechraten mit arbeitsbezogenen Endpunkten konnten wegen unzureichender Anzahl verfügbarer Studien nicht durchgeführt werden.

3.7.1 Regressionen der schmerzbezogenen 30%-Ansprechraten mit den standardisierten, arbeitsbezogenen Endpunkten

Es wurden jeweils für die relative (RR) und die absolute (RD) Berichterstattungsweise der schmerzbezogenen 30%-Ansprechraten lineare Regressionen mit den standardisierten Ergebnissen der Beeinträchtigung bei der Arbeit erstellt (Abbildung 33).

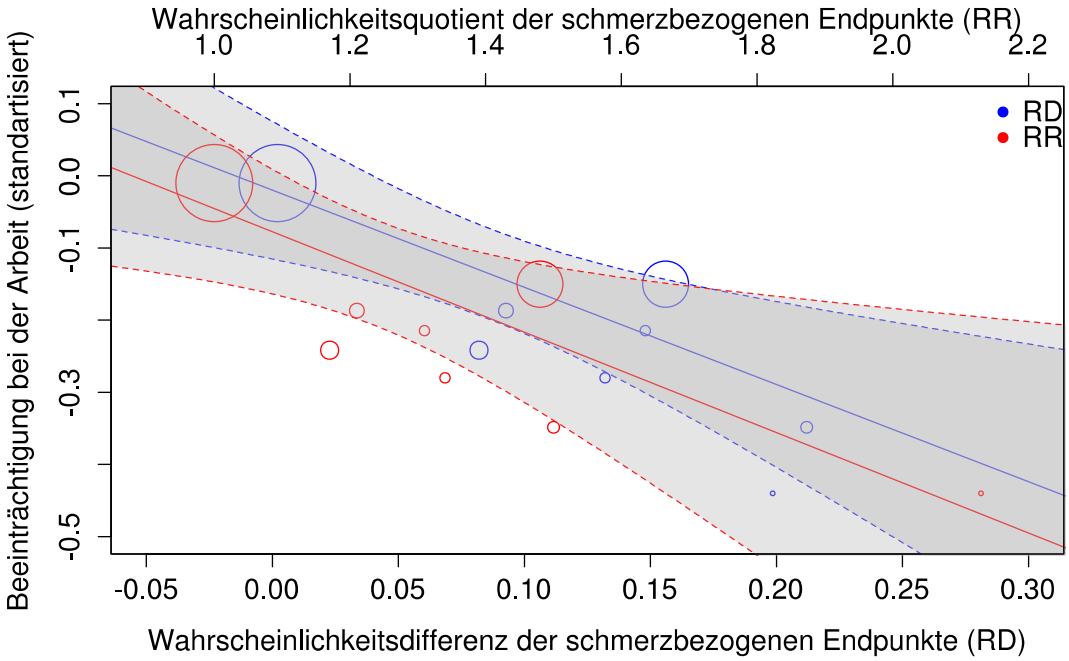


Abbildung 33: *Regressionen der Wahrscheinlichkeitsdifferenz (RD) und des Wahrscheinlichkeitsquotienten (RR) von schmerzbezogenen 30%-Ansprechraten mit der standardisierten Beeinträchtigung bei der Arbeit.*
Die Größe der Punkte repräsentiert die individuelle Gewichtung der einzelnen Studien basierend auf ihrer inversen Varianz.
Die gestrichelten Linien markieren das 95%-Konfidenzintervall.

Die Differenzen (RD) der 30%-Ansprechraten zwischen aktiver und Placebobehandlung zeigen eine statistisch signifikante (p=0,008) Relation mit den

standardisierten Mittelwertsdifferenzen der kontinuierlichen arbeitsbezogenen Endpunkten. Die Steigung beträgt -1,34.

Für den RR ließ sich ebenfalls ein linearer Zusammenhang nachweisen (Steigung -0,37), welcher mit p=0,021 statistische Signifikanz erreicht.

3.7.2 Regressionen der schmerzbezogenen 50%-Ansprechraten mit den standardisierten, arbeitsbezogenen Endpunkten

Die Wahrscheinlichkeitsquotienten und die Wahrscheinlichkeitsdifferenzen der schmerzbezogenen 50%-Ansprechraten wurden mithilfe lineare Regressionen in Relation zu den standardisierten Ergebnissen der Beeinträchtigung bei der Arbeit gesetzt (Abbildung 34).

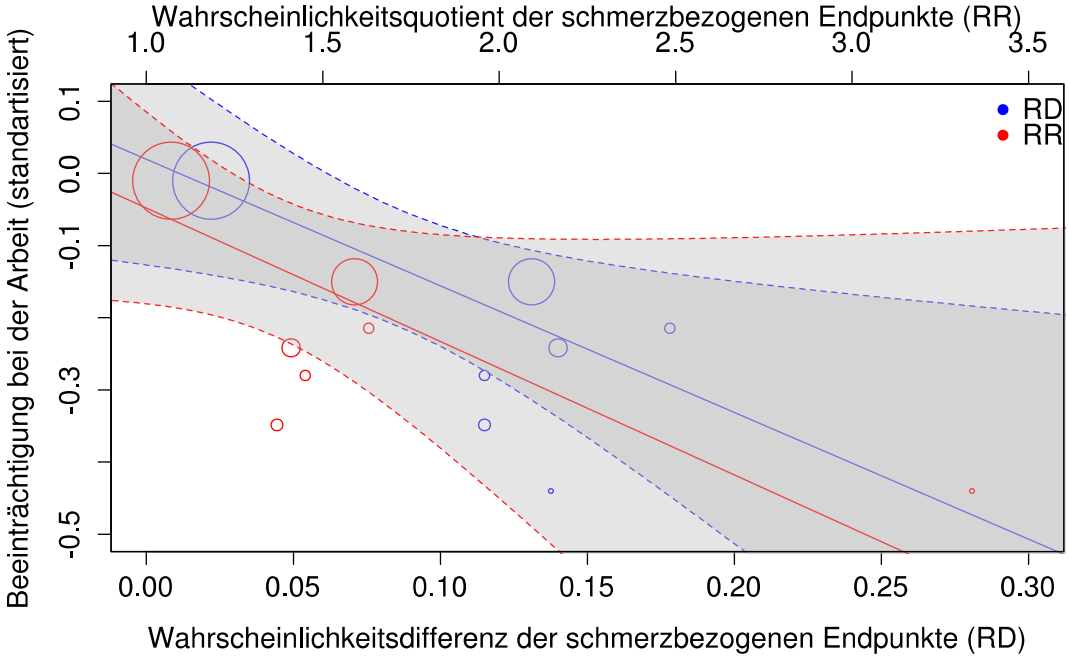


Abbildung 34: *Regressionen der Wahrscheinlichkeitsdifferenz (RD) und des Wahrscheinlichkeitsquotienten (RR) von schmerzbezogenen 50%-Ansprechraten mit der standardisierten Beeinträchtigung bei der Arbeit.*
Die Größe der Punkte repräsentiert die individuelle Gewichtung der einzelnen Studien basierend auf ihrer inversen Varianz.
Die gestrichelten Linien markieren das 95%-Konfidenzintervall.

Für die Differenz der Wahrscheinlichkeiten (RD) war ein signifikanter (p=0,025) Zusammenhang mit der Steigung -1,75 nachweisbar. Die Steigung der Regressionsgeraden der relativen Wahrscheinlichkeiten (RR) beträgt -0.22 und verfehlt mit einem p-Wert von 0.065 knapp die Signifikanzschwelle.

3.7.3 Heterogenitäts- und Sensitivitätsanalysen zu den Abschnitten 3.7.1 und 3.7.2

Es zeigte sich, dass höhere Werte von RR und RD bezogen auf die Schmerzendpunkte auch zu analogen Verbesserungen hinsichtlich der arbeitsbezogenen Endpunkte führten. Einschränkend ist zu berücksichtigen, dass die Meta-Analyse von Straube et al. 2011 48% der Patienten stellt und gleichzeitig in ihren Ergebnissen stark von den der anderen Studien abweicht. Dies könnte Ursache für eine mögliche Verzerrung der Regression sein. In Sensitivitäts-analysen ohne Einbeziehung der Meta-Analyse von Straube et al. 2011 konnten keine statistisch signifikanten Trends nachgewiesen werden.
Zusammenfassend lässt sich feststellen, dass es möglich war einen linearen Zusammenhang zwischen den Ansprechraten schmerzbezogener Endpunkte und den standardisierten arbeitsbezogenen Endpunkten nachzuweisen.

4. Diskussion

4.1 Einschränkungen der verwendeten Modelle und Verfahren

4.1.1 Vollständigkeit der Literaturrecherche

Trotz des hohen Aufwands für die Entwicklung und Durchführung der Suche (Abschnitt 2.1), unterliegt diese dennoch einer Reihe nicht vernachlässigbarer Einschränkungen. So ist die Beschränkung auf englische und deutsche Artikel den linguistischen Fähigkeiten des Autors zuzurechnen. Die verwendete Datenbank, MEDLINE®/PubMED®, stellt das wichtigste Verzeichnis medizinischer Artikel dar, erfasst aber nicht vollständig alle publizierten wissenschaftlichen Zeitschriften. Weiterhin ist es möglich, dass in relevanten Publikationen ungebräuchliche Begriffe verwendet wurden, oder dass Studien bei der Erfassung durch die Datenbank gar nicht oder mit unpassenden Kategorien verknüpft wurden und daher nicht durch die in dieser Arbeit verwendeten Suchwörter identifiziert wurden. Daher wurde, um die Möglichkeit eines systematischen Fehlers aufgrund fehlerhafter Datenbankverknüpfungen zu minimieren, die Datenbankabfrage ohne die von PubMED® angebotenen Filter durchgeführt.

4.1.2 Studienqualität

Da ein Review nur die veröffentlichte Datenlage abbilden kann, ist zur Einordnung der Ergebnisse die Bewertung der Studienqualität sehr wichtig.

Hierzu wurde die OQS das nach Olivo et al. 2008 meist zitierte und das meist verwendete Untersuchungswerkzeug im Bereich der medizinischen Arbeiten genutzt.
Dieses etablierte Instrument wurde mit der Delphi-Methode entwickelt. Bei der Delphi-Methode werden Experten der entsprechenden Fachrichtungen in einem

mehrstufigen, möglichst anonymen Verfahren um Vorschläge und gegenseitige Bewertungen der Vorschläge gebeten, um so durch jede Stufe die Entscheidungsfindung zu verfeinern. Es soll den Experten somit möglich sein, durch Rückmeldungen und gegenseitige Inspirationen, aber ohne gruppen-dynamische Einflüsse, eine systematische und kollektive Aussage zu treffen (Brown 1968).

Das Instrument OQS erschien wegen seiner Einfachheit und vor dem Hintergrund, dass es im Bereich von Schmerzstudien entstand, als am besten geeignet zur Bewertung der Studienqualität für die vorliegende Dissertation. Dennoch sind einige Einschränkungen zu nennen. Eine Stärke des Systems ist seine Einfachheit. Es werden nur drei Eigenschaften bewertet, während andere Instrumente leicht über zehn Punkte abgearbeitet sehen wollen. Diese Einfachheit bringt eine recht schnelle Bewertung und eine gewisse Homogenität der Ergebnisse bei Reevaluierung mit sich. So berichteten Olivo et al. 2008 von einer Korrelation von bis zu 1 zwischen verschiedenen Bewertern und Evaluierungen, was als perfektes Ergebnis betrachtet werden kann. Allerdings führt die Reduktion auf nur drei Gesichtspunkte, die teilweise bereits durch die Erwähnung eines Wortes mit einem Bewertungspunkt honoriert werden, zwingend zu einem Informationsverlust. So erhält eine Studie bereits 2 Punkte, wenn sie erwähnt, dass randomisiert wurde und die Behandlungen doppelblind erfolgten. Im weiteren Verlauf der Bewertung können diese Punkte wieder entzogen werden, sollte sich herausstellen, dass die dabei verwendeten Techniken nicht angebracht waren. Wird jedoch nichts Weiteres berichtet, so behält die Studie ihre zwei Punkte. Dies bedeutet, dass fehlende Berichterstattung zu einer möglicherweise nicht gerechtfertigten hohen Bewertung führen kann. Gleichzeitig könnte eine korrekt durchgeführte Studie unterbewertet werden, da sie die Details der Durchführung verschweigt. Die Lösung der OQS-Entwickler, die Teilfragen mit den übertragbaren Kategorien „inkorrekt“ (0 Punkte), „keine Details veröffentlicht“ (1 Punkt) und „korrekt durchgeführt und berichtet“ (2 Punkte) zu bewerten, versucht als Kompromiss den Platzbeschränkungen und den damit verlustbehafteten, komprimierten Berichterstattungen in den klassischen Publikationsmedien Rechnung zu tragen. Wie schon Jadad et al. 1996 richtig bemerken, leidet die

Aussagekraft des Instrumentes unter mangelnder oder ungenauer Berichterstattung.
Als weitere Einschränkung fiel eine Unterbewertung des Gesichtspunkts der Studienabbrecher auf. Nahezu keine größer angelegte Studie kann ohne fehlende Messwerte beendet werden. Ganz gleich, ob diese durch Fehler der Durchführenden oder durch vorzeitiges Abbrechen der Probanden verursacht werden, ist es für die Interpretation der Ergebnisse wichtig, den statistischen Umgang mit den fehlenden Daten zu beschreiben. Auf der OQS kann diesbezüglich ein Punkt vergeben werden. Dieser honoriert die Erwähnung von Studienabbrechern und deren Grund für das Ausscheiden. Die weitere statistische Behandlung der daraus folgenden lückenhaften Daten wird jedoch nicht weiter untersucht und bewertet. So ist eine Dokumentation der statistischen Methoden, oder gar eine Intention-To-Treat-Analyse (ITT) nicht gefordert.

Abhängig von dem Studiendesign und der Anzahl der Dropouts ist eine genaue Darstellung des statistischen Umgangs mit fehlenden Daten für die Nachvollziehbarkeit der Ergebnisse erforderlich. Wie die Übersichtsarbeit von Hollis und Campbell 1990 zeigte, kann und sollte die einfach erscheinende Maßgabe von ITT, nämlich die individuellen Patientendaten in den Gruppen zu analysieren, denen sie zu Beginn der Studie zugeteilt wurden, auf verschiedene Weise durchgeführt werden. Weiterhin zeigen sie, dass fehlende Daten keine Seltenheit sind. Fast ein Viertel der Studien wiesen über 10% fehlendes Erfassen eines Endpunktes auf. Um eine Verzerrung der Ergebnisse zu verringern, werden verschiedene Herangehensweisen mit ihren jeweiligen Grundannahmen herangezogen (Wood et al. 2004). So können unvollständige Daten ganz aus der Analyse entfernt werden, wenn die Grundannahme erfüllt ist, dass der Abbruch der Datenerfassung nicht auf die Behandlung zurückzuführen ist. Diese Voraussetzung kann jedoch in der Regel nicht unterstellt werden. Alternativ können die letzten verfügbaren Daten (*last observation carried forward*, LOCF) oder die Werte zu Studienbeginn (*baseline observation carried forward*, BOCF) für die Analyse verwendet werden. Darüber hinaus können Regressionen zum Einsatz kommen, um die fehlenden Daten aus den bereits vorhandenen zu schätzen (Wood et al. 2004). Wurden wiederholt Daten zu verschiedenen

Zeitpunkten erhoben, so können bei einem *mixed model for repeated measurements* (MMRM) Kovarianzanalysen für jeden Untersuchungszeitpunkt erstellt werden (Lane 2008).

Das Schätzen von fehlenden Daten aufgrund nicht zu überprüfender Grundannahmen kann jedoch problematisch sein. So zeigte Lane 2008, dass LOCF oft nicht in Richtung einer Effektverringerung verzerrt. Gerade wenn die Ausfallrate in der Placebogruppe deutlich höher ist als in der aktiven Gruppe, kann es auch zu einer Überbewertung des Behandlungseffektes kommen. Moore et al. kamen 2012b ebenfalls zu dem Schluss, dass gerade im Zusammenhang mit Ansprechraten LOCF, verglichen mit BOCF, Behandlungseffekte der aktiven Gruppe überschätzt. Daher ist bei Analysen, bei denen nur Baseline und Endpunktdaten erhoben wurden, BOCF als konservativere Schätzung der LOCF-Methodik vorzuziehen. Es wird jedoch auch bei BOCF ebenfalls mit Daten gerechnet, die ursprünglich nicht erhoben wurden und es ergibt sich damit immer die Gefahr, unvorhersehbare Verzerrungen zu verursachen. Lane forderte daher 2008 eine allgemein häufigere Verwendung eines *mixed model for repeated measurements* (MMRM) bei dem keine nicht vorhandenen Daten geschätzt werden, sondern eine Kovarianzanalyse der vorhandenen Daten durchgeführt wird. Dieses Design setzt jedoch eine mehrfache Wiederholung der Messungen über den Verlauf der Studie hinweg voraus und erhöht damit deutlich den klinischen und finanziellen Aufwand.

Von den 31 in das systematische Review eingeschlossenen Publikationen erhoben sieben für ihre primären Endpunkte[7] wiederholt Daten, die dann mit einem MMRM analysiert wurden (Arnold et al. 2005, Arnold et al. 2009, Chappell et al. 2009, Chappell et al. 2011, Skljarevski et al. 2009, Skljarevski et al. 2010a, Skljarevski et al. 2010b). Für ihre sekundären Endpunkte, die auch die arbeitsbezogenen Endpunkte einschlossen, wurde jedoch nur eine LOCF Analyse durchgeführt. Tatsächlich war das LOCF-Vorgehen die häufigste Form, mit fehlenden Daten umzugehen. Insgesamt 16 Arbeiten berichteten diese Herangehensweise (Arnold et al. 2005, Arnold et al. 2009, Bradley et al. 2010,

7 Primärer Endpunkt: der wichtigste untersuchte Endpunkt

Carlsson und Sjölund 2001, Chappell et al. 2008, Chappell et al. 2009, Chappell et al. 2011, Kavanaugh et al. 2009a, Kavanaugh et al. 2013, Manchikanti et al. 2010, Markenson et al. 2005, Skljarevski et al. 2009, Skljarevski et al. 2010a, Skljarevski et al. 2010b, Smolen et al. 2006, Strand et al. 1999). Elf Publikationen berichteten gar keinen statistischen Umgang mit fehlenden Daten, oder die berichteten Vorgehensweisen konnten nicht mit den präsentierten Ergebnissen in Zusammenhang gebracht werden (Albert et al. 2013, Baron et al. 2010, Bejarano et al. 2008, Bennett et al. 2003 & 2005, Egsmose et al. 1997, Kavanaugh et al. 2006, Lehmann et al. 1986, Licciardone et al. 2003, Meireles et al. 2010, van der Heijde et al. 2006).

Nur zwei Arbeiten verwendeten zusätzlich zu LOCF die BOCF-Methodik (Chappell et al. 2011, Skljarevski et al. 2010b). Bei zwei kleineren Studien gab es keine fehlenden Daten (Barkham et al. 2010, He et al. 2005) und eine Studie untersuchte nur die Daten von Patienten, die bis zum Ende Teil der Studie waren (Jarzem et al. 2005). In dem Review von Straube et al. 2011 wurden aufgrund der unterschiedlichen Herangehensweisen in den Primärpublikationen Daten aus BOCF- und LOCF-Analysen kombiniert.

Dies zeigt, dass arbeitsbezogene Endpunkte, selbst wenn sie in großen Studien untersucht und berichtet werden, eher eine untergeordnete Rolle spielen. So wurden die in diesen Studien zu arbeitsbezogenen Endpunkten erhobenen Daten gegenüber den primären Endpunkten nur mit einfachen statistischen Analysen ausgewertet.

Ungeachtet dessen wurden für diese Dissertation viele hochqualitative Arbeiten mit zahlreichen Probanden verwendet. Von den 27 erfassten Patientenkollektiven untersuchten nur 13 weniger als 100 Probanden in jedem Behandlungsarm (Albert et al. 2013, Barkham et al. 2010, Bejarano et al. 2008, Carlsson und Sjölund 2001, Egsmose et al. 1997, He et al. 2005, Jarzem et al. 2005, Lehmann et al. 1986, Licciardone et al. 2003, Manchikanti et al. 2010, Markenson et al. 2005, Meireles et al. 2010, van der Heijde et al. 2006).

Wie Zhang und Xu 2013 zeigten, weisen diese als klein definierten Studien durchschnittlich größere Behandlungseffekte auf als größer angelegte Studien.

Sie führen dies auf Publikationsverzerrungen und geringere Studienqualität bei kleineren Forschungsprojekten zurück. Für das vorgelegte Review bedeutet dies jedoch auch im Umkehrschluss, dass etwas über die Hälfte der analysierten Studien angemessen groß waren. Von den insgesamt 11434 erfassten Patienten tragen die kleinen Studien nur 1593, also etwa 14% bei, während die großen Studien 9841 Patienten (86%) untersuchten. Wie auch schon in anderen Übersichtsarbeiten, bei denen große Diskrepanzen der Patientenzahlen zwischen den untersuchten Studien auftraten (Straube et al. 2009), konnte beobachtet werden, dass aufgrund des geringen Gewichtes, welches die kleinen Studien erhalten, die Kernaussage der Meta-Analysen auf den Ergebnissen der großen Studien beruht. Hätte man in der Analyse von Abschnitt 3.4.1.1 die kleinen Studien ausgeschlossen, so wäre das Ergebnis nur geringfügig anders gewesen. So ergab sich in der Analyse mit allen Studien eine durchschnittliche standardisierte Mittelwertsdifferenz von -0,35 (95%-KI: -0,50; -0,21). Demgegenüber betrug ohne die Studien mit unter 100 Probanden pro Behandlungsarm (Barkham et. 2010, Bejarano et al. 2008, Markenson et al 2005, van der Heijde et al. 2006) die durchschnittliche standardisierte Mittelwertsdifferenz -0,34 (95%-KI: -0,50; -0,17). Die Aussagekraft einer Meta-Analyse scheint also grundlegend davon abzuhängen, ob Daten von großen Studien verfügbar sind.

4.1.3 Datenextraktion aus Interquartilangaben

Für die Meta-Analysen in den Abschnitten 3.4.1, 3.4.5 und 3.5.1 mussten die verwendeten Daten aus den Studien Albert et al. 2013, Kavanaugh et al. 2006 und van der Heijde et al. 2006 anhand von Interquartilangaben geschätzt werden. Dazu wurde angenommen, dass die individuellen Patientendaten normalverteilt waren. Für den Durchschnittseffekt wurde der Median als Schätzer verwendet und die Standardabweichung wurde mit einem Umrechnungsfaktor aus der Interquartilbreite geschätzt (Abschnitt 2.3). Da Autoren diese Art der Datenpräsentation oft wählen, wenn die Daten nicht normalverteilt sind, wurden,

wie von Higgins und Green 2011 vorgeschlagen, Sensitivitäts-Analysen erstellt, um den Einfluss der geschätzten Daten in den jeweiligen Meta-Analysen zu untersuchen:

- In der durchgeführten Meta-Analyse aller verfügbaren Studien (Abschnitt 3.4.1.1) zeigte sich, dass die, wie zuvor beschrieben, aus Interquartilangaben geschätzten Mittelwertsdifferenzen von Kavanaugh et al. 2006 ähnlich stark von den Ergebnissen der restlichen Studien abwichen wie die vollständig berichteten Ergebnisse von Kavanaugh et al. 2013, sie jedoch auch die Heterogenität erhöhten. Eine erstellte Sensitivitätsanalyse ohne die Studien von Kavanaugh et al. 2006, Kavanaugh et al. 2013 und Straube et al. 2011 ergab einen um 0,07 schwächeren standardisierten Behandlungseffekt.

- In Abschnitt 3.4.5.1 erhöhten die aus Interquartilangaben geschätzten Daten der Studien Kavanaugh et al. 2006 und van der Heijde et al. 2006 nicht die Gesamtheterogenität, sondern senkten mit ihren weiten Konfidenzintervallen den ursprünglichen Wert von I^2 von 92% auf 82%. Der standardisierte Behandlungseffekt wurde durch sie um 0.10 erhöht, das Konfidenzintervall konnte deutlich eingeengt werden.

- In den erstellten Meta-Analysen in Abschnitt 3.5.1 wichen die Ergebnisse aus Interquartilangaben von Albert et al. 2013 und van der Heijde et al. 2006 teils erheblich von denen der anderen Studien ab. Bei der Analyse der Mittelwertsdifferenzen erhöhte sich durch die Einbeziehung der Daten von van der Heijde et al. 2006 der gepoolte Behandlungseffekt um 0,09.

4.2 Heterogenität

4.2.1 Kontinuierliche Endpunkte

4.2.1.1 Standardisierung der Studienergebnisse

Die vielfältigen Erscheinungsformen von arbeitsbezogenen Endpunkten führen zu einer ebenso vielfältigen Berichterstattung. Wie in Abschnitt 3.3 beschrieben wurden in den Studien insgesamt 16 verschiedene Endpunkte untersucht, die wiederum heterogen berichtet wurden. Idealerweise sollte für die Zusammenfassung der Daten in einer Meta-Analyse jedoch nur ein Endpunkt nach einheitlichen Gesichtspunkten beschrieben werden. So ist es beispielsweise nicht möglich, Daten verschiedener Instrumente zu kombinieren, da eine gemessene Veränderung auf der Skala des einen eine völlig andere Effektamplitude bedeuten könnte als auf der eines anderen. Für viele der erstellten Meta-Analysen mussten die untersuchten Studienergebnisse daher zunächst standardisiert werden. Die hierfür erforderliche Grundannahme war, dass alle einfließenden Messinstrumente einen gemeinsamen zu Grunde liegenden Endpunkt messen sollten. Dies bedeutet, dass die Standardabweichung eines Behandlungseffektes über alle Studien hinweg gleich ist, und dass die Unterschiede des Wertes allein durch die unterschiedlichen Messinstrumente oder Einheiten bedingt sind (Higgins und Green 2011). Variationen des Effektes sowie dessen Standardabweichung aufgrund von unterschiedlichen Behandlungen oder Studienpopulationen werden so ignoriert. Nach Cummings 2011 sind diese Annahmen jedoch oft nicht gerechtfertigt oder lassen sich in Ermangelung des gesamten Datensatzes nicht überprüfen.

Eine weitere Einschränkung der Standardisierung ist die schwierige Interpretation der als Vielfaches der Standardabweichung ausgedrückten Ergebnisse (standardisierte Mittelwertsdifferenz, SMD). Hierzu verwendete Cohen 1988 als erster eine vergleichbare Standardisierung und gab einen Interpretationsschlüssel an, bei dem 0,2 einen kleinen, 0,5 einen mittleren und 0,8 einen großen

Behandlungseffekt beschreiben. Cummings 2011 wertet diese Einteilung jedoch als problematisch, da sie willkürlich von Cohen definiert sei.

4.2.1.2 Transformierung der SMD in klinische Endpunkte

Higgins und Green schlagen 2011 vor die SMD in eingängige Instrumente zu transformieren. So könnte man die SMD mit für die untersuchte Population und den Endpunkt typischen Standardabweichungen auf das ursprünglich untersuchte Messinstrument normieren. Diese Vorgehensweise geht von der Annahme aus, dass die generelle, über alle Studien gleiche Standardabweichung zwischen den Probanden bekannt wäre. Da dies selten der Fall ist, schlugen Higgins und Green 2011 vor, Interpersonen-Standardabweichungen für diese Zwecke aus repräsentativen empirischen Studien zu entnehmen. In Ermangelung empirischer Studien wurden in dieser Arbeit die von den Studien berichteten Baselinewerte als Schätzer der generellen Standardabweichung verwendet. Dabei ist es möglich, dass bei Auswertung eventuelle Unterschiede in der rekrutierten Probandenpopulation zu vermeintlich unterschiedlichen Behandlungseffekten führen. Die in dieser Arbeit erstellten Umrechnungen der SMD in Beeinträchtigungspunkte des BPI (Abschnitt 3.4.3.1) und in VAS-Bereiche (Abschnitt 3.4.5) sollen daher nur als interne Konsistenzkontrollen und als Orientierungshilfe für den Leser dienen:

- Die Ergebnisse der Umrechnung zeigen eindringlich die Möglichkeiten und Einschränkungen der verwendeten Methoden. So kam eine Umrechnung der in Abschnitt 3.4.1.1 errechneten SMD-Werte in Beeinträchtigungspunkte des BPI und des FIQ (Abschnitt 3.4.3.1) zu fast derselben Mittelwertsdifferenz wie die Meta-Analyse der Studien, die arbeitsbezogene Endpunkte mithilfe von BPI und FIQ untersuchten (Abschnitt 3.4.3), wies aber ein etwas größeres Konfidenzintervall auf.

- In Abschnitt 3.4.5.3 wurde der gleiche SMD (Abschnitt 3.4.1.1) in eine visuelle Analogskala für Arbeitsproduktivität umgerechnet. Hier zeigte sich,

dass der aus der Umrechnung des SMD entstandene geschätzte Effekt deutlich geringer war als der, welcher sich auf Berechnungen der direkt gepoolten VAS-Ergebnisse (Abschnitt 3.4.5.2) gründete. Dies lässt sich mit der zahlenmäßigen Dominanz der Studien erklären, die das BPI oder das FIQ erhoben hatten. So bekamen die 3 Studien, die eine VAS verwendeten, nur auf 21,6% der kumulativen Gewichtung in der Meta-Analyse von Abschnitt 3.4.1.1.

Die zwei genannten Beispiele zeigen, dass die Umrechnung des SMD in greifbare Messskalen mitunter zur Illustration und Einordnung dienen kann, aber aufgrund möglicher Fehlinterpretationen nicht isoliert als abschließendes Ergebnis berichtet werden sollte.

4.2.1.3 Transformierung der SMD in den odds ratio

Chinn schlug 2000 vor das Quotenverhältnis, den *odds ratio* (OR), in die SMD umzurechnen, um so Studien mit kontinuierlichen und binären Endpunkten kombinieren zu können. Im Umkehrschluss könnte man auf diese Weise auch die SMD in das OR umwandeln. Dies ist allerdings nur möglich, wenn es einen zu Grunde liegenden kontinuierlichen Endpunkt gibt. Dabei wird argumentiert, dass der oftmals willkürlich gesetzte Schwellenwert, der einen kontinuierlichen Endpunkt zu einem dichotomen Endpunkt definiert, keinen Einfluss auf das OR hat, wenn der Endpunkt einer logistischen Verteilung mit gleicher Varianz in beiden Behandlungsgruppen folgt. Da sich die Normalverteilungsfunktion und die logistische Funktion hauptsächlich in ihren Endverläufen unterscheiden, kann als Näherung der Normalverteilung die logistische Verteilung eingesetzt werden. Die standardisierte logistische Verteilung hat eine Varianz von $\pi^2/3$ und soll als Rechenfaktor für die Transformierung dienen. Die Umrechnung erfolgt entsprechend Formel 5.

$$\ln \text{ OR} = \frac{\pi^2}{3} \times SMD$$

$\text{OR}: Odds\ Ratio$
$SMD: standardisierte\ Mittelwertsdifferenz$

Formel 5

Mit dem so erhaltenen OR könnte man mit bekannten Ansprechraten der Kontrollgruppen einen Number-Need-To-Treat-Wert (NNT) schätzen, der für klinische Szenarien möglicherweise besser interpretierbar wäre (Higgins und Green 2011). Es ergibt sich jedoch ein ähnliches Problem wie schon zuvor bei der Umrechnung der SMD in bekannte Messinstrumente (Abschnitt 4.2.1.2). Wie bereits erwähnt, wurden arbeitsbezogene Endpunkte in der Literatur selten untersucht. Es gibt daher keine repräsentative Studie, die eine zu erwartende Ansprechrate für arbeitsbezogene Endpunkte bezüglich der Kontrollgruppe berichtet. Wird statt dessen eine willkürlich gewählte Ansprechrate für einen speziellen Endpunkt verwendet, so könnte es zu Verzerrungen kommen. Das Ergebnis könnte hier weder in Anspruch nehmen den speziellen Endpunkt, noch irgendeinen anderen, realistisch abzubilden. Weiterhin könnte ein Ausdrücken der SMD in einem vertrauten Messinstrument oder als NNT das Verständnis für die Fragestellung der Arbeit erschweren. Nicht nur die Heterogenität der Endpunkte stellt sich als Herausforderung dar: Bereits die allgemeine Fragestellung, die verschiedene Erkrankungen und Behandlungen einschließt, macht eine klinische Interpretation der Ergebnisse hinfällig.

Die Einschränkungen der SMD als Berichterstattung erschweren Aussagen bezüglich Fragestellungen nach spezifischen Ergebnissen. Sie spiegeln damit die Datenlage wider. Die berichteten Effekte, ausgedrückt durch standardisierte Mittelwertsdifferenzen, ermöglichen es, eine grobe Richtungsangabe über den allgemeinen Stand des Behandlungserfolges bezüglich arbeitsbezogener Endpunkte zu geben. Eine Prognose auf klinischer Ebene für individuelle Patienten ist damit nicht möglich.

4.2.2 Ergebnisinterpretation

Aufgrund der in den vorherigen Abschnitten dargestellten Einschränkungen war es bei der Erstellung dieser Arbeit wiederholt nötig, parallele Berechnungen mit verschiedenen Datenkombinationen durchzuführen. Durch diese Sensitivitätsanalysen war es möglich, den Einfluss einzelner Studienergebnisse auf die Gesamtanalyse zu untersuchen. Ein wichtiger Indikator für die Heterogenität war hier der bereits erwähnte I^2-Wert (Abschnitt 2.4.1). Gerade bei der ersten großen Meta-Analyse (Abschnitt 3.4.1.1) fand sich ein recht hoher Wert von 81%. Wie Higgins et al. 2003 feststellten, spricht dies dafür, dass ein Großteil der Inkonsistenz auf echter Heterogenität und nicht auf Zufall beruht. Daher wurde die Ursache für die Heterogenität näher untersucht. Wie die Regressionsanalyse in Abschnitt 3.4.1.2 zeigt, gibt es Hinweise, dass die verschiedenen Einschätzungen des Behandlungseffektes mit unterschiedlicher Studienqualität assoziiert sind. Hierbei konnte ein statistisch signifikanter Zusammenhang zwischen der mit der OQS bewerteten Studienqualität und den jeweiligen Studienergebnissen nachgewiesen werden.

Gleichzeitig wurden zur Erstellung der Meta-Analysen zwei verschiedene Modelle an den Datensätzen erprobt (Abschnitt 3.4.1.2). Bei dem einen Modell wurde davon ausgegangen, dass alle Studien einen allgemeingültigen, in allen Studienpopulationen gleichartigen Effekt messen (*fixed effect model*, FE). Dazu wurden die Studien mit abweichenden Effektamplituden (Kavanaugh et al. 2006, Kavanaugh et al. 2013, Straube et al. 2011) entfernt. Bei dem anderen Modell wurde angenommen, dass in jeder Studie ein anderer Behandlungseffekt vorhanden war, welcher mit der Varianz Tau^2 um einen Durchschnittswert streut (*random effects model*, RE). Hierbei wurden die Daten aller Studien verwendet. Beide Analysen kamen zu ähnlichen Werten (FE 0,28; RE 0,35). Die Wahl des RE für die finale Präsentation schien aufgrund des weiteren Konfidenzintervalles und dem Fakt, dass die Ergebnisse aller Studien einfließen konnten als die konservativere Möglichkeit.

Bei den folgenden Meta-Analysen in den Abschnitten 3.4.2 bis 3.4.6 waren die Einschlusskriterien für die auszuwertenden Studien enger gesteckt. Für die durchgeführten Meta-Analysen von spezifischen arbeitsbezogenen Endpunkten waren daher die I^2-Werte geringer.

Trotz bestehender Unterschiede zwischen den Ergebnissen der untersuchten Studien ergab sich ein deutliches und recht einheitliches Bild vom Behandlungseffekt kontinuierlicher arbeitsbezogener Endpunkte. Die standardisierten Mittelwertsdifferenzen der einzelnen für bestimmte arbeitsbezogene Endpunkte spezifischen Analysen (Abschnitte 3.4.2 bis 3.4.6) schwankten zwischen ca. 0,2 und 0,5. Das Ergebnis der Meta-Analyse, die alle Studien mit arbeitsbezogenen Endpunkten mit einschloss (Abschnitt 3.4.1.1), reihte sich in diesen Ergebnisbereich mit einem standardisierten Effekt von etwa 0,3 mittig ein.

4.2.3 Ansprechraten

Bei den Meta-Analysen der Ansprechraten (Abschnitt 3.6) zeigte sich erneut das Heterogenitätsproblem. Die Diskrepanzen der Schmerzansprechraten konnten auf das Verwenden von verschiedenen Therapeutika als mögliche Ursache zurück geführt werden (Abschnitt 3.6.1.3). Jedoch können auch Einflüsse der untersuchten Studienpopulationen oder der Studiendurchführung nicht ausgeschlossen werden. Da man bei Literaturrecherchen auf die vorhandene Berichterstattung angewiesen ist, sind letztere Faktoren sehr schwer zu untersuchen.

Die Analysen der ACR-Ansprechraten (Abschnitt 3.6.2.4) zeigten einen deutlichen Unterschied zwischen relativer und absoluter Berichterstattung auf. Wie dort bereits beschrieben, waren die Varianzen der absoluten Ergebnispräsentation (RD) zwischen den zu Grunde liegenden Behandlungseffekten der in den Meta-Analysen verwendeten Studien weitaus geringer als die der relativen (RR), obwohl sie auf den gleichen Daten beruhten. Dies verwunderte, da in der Literatur ein gegenteiliger Effekt dokumentiert wurde. So stellte Deeks 2002 in seiner

Analyse von 551 systematischen Reviews fest, dass relative Berichterstattungen wie *odds ratio* (Quotenverhältnis) oder *relative risk* (Wahrscheinlichkeitsquotienten) homogenere Ergebnisse lieferten als die absoluten Wahrscheinlichkeitsdifferenzen (*risk difference*). Mit den im Abschnitt 3.6.2.4 erstellen Regressionen konnte ein statistisch signifikanter Zusammenhang zwischen den verschiedenen Ergebnissen der Kontrollgruppen und den Ergebnissen der entsprechenden Studien nachgewiesen werden. Der Einfluss der Kontrollgruppenergebnisse auf das Gesamtergebnis einer Studie kann daher durchaus als Grund für bestehende Heterogenität zwischen den Studien angesehen werden. Weiterhin zeigten diese Regressionen, dass sowohl die absolute als auch die relative Präsentationsform für Verzerrungen durch verschiedene Ansprechraten innerhalb der jeweiligen Kontrollgruppen anfällig waren. Mithilfe der vorliegenden Ergebnisse kann die Frage nach den Gründen der scheinbar erhöhten Robustheit gegenüber Verzerrungen der absoluten Präsentationsform derzeit nicht beantwortet werden.

4.3 Arbeitsbezogene Endpunkte in der Literatur

Wie in der Einleitung beschrieben (Kapitel 1), verursachen chronische Schmerzerkrankungen jährlich beträchtliche wirtschaftliche Schäden und schränken den professionellen und privaten Alltag eines großen Teiles der Gesellschaft stark ein. Das Interesse der wissenschaftlichen Gemeinschaft, arbeitsrelevante Endpunkte in randomisierten, kontrollierten Studien zu diskutieren, ist bisher jedoch gering (Abschnitt 3.2). Nur 0,23% aller Publikationen, die die Behandlung chronischer Schmerzen in randomisierten, placebokontrollierten Studien zum Thema hatten, beschrieben arbeitsbezogene Endpunkte. Das Interesse an diesem für das Individuum und sein Umfeld wichtigen Aspekt der Erkrankungen scheint seit Mitte der 1990er Jahre geweckt zu sein (Abschnitt 3.2), blieb aber relativ zum Forschungsaufwand die letzten 20 Jahre hinweg gering. Wie in Abschnitt 4.1.2 erläutert, zeugen auch der statistische

Umgang und die Art der Präsentation der Daten von der sekundären Bedeutung der arbeitsbezogenen Endpunkte.

Oft sind die Ursachen und Mechanismen der einzelnen Schmerzerkrankungen noch nicht ausreichend aufgeklärt und entziehen sich einer zufriedenstellenden Behandlung. Weiterhin sind Schmerzen im rheumatischen Bereich häufig mit autoimmunologischen und degenerativen Prozessen verbunden, deren Symptome und Verlauf nur abgemildert oder aufgehalten werden können. Eine Regeneration und damit eine Entfernung der Ursache der Schmerzen ist jedoch kaum möglich. Daher wird versucht, den Schmerz als das zu behandelnde Hauptsymptom zu verringern, mit der Erwartung, auch die arbeitsrelevanten Einschränkungen zu beeinflussen. Diese Annahme scheint gerechtfertigt zu sein, da in den Abschnitten 3.5.2 und 3.7 gezeigt werden konnte, dass die schmerzbezogenen und arbeitsbezogenen Endpunkte einen linearen Zusammenhang aufweisen. Gleichzeitig gibt es Hinweise (Carlier et al. 2013), dass umgekehrt auch der Beschäftigungsstatus und die damit verbundenen psychischen und physischen Verhaltensprofile den selbst empfundenen Gesundheitszustand beeinflussen.

Eine weitere Ursache der geringen Beachtung von arbeitsbezogenen Endpunkten in der Literatur könnte in der Komplexität dieser Endpunkte zu suchen sein. Zur Komplexität der arbeitsbezogenen Endpunkte trägt auch die Tatsache bei, dass Schmerzen nicht die einzigen Limitationen sind, die „Schmerz“-Patienten in Kauf nehmen müssen. So sind bei vielen entzündlichen, rheumatischen Erkrankungen Gelenkdegenerationen die Ursache für teilweise umfangreiche Funktionseinschränkungen (Hettenkofer et al. 2003). Selbst Krankheiten, die als klassische Schmerzerkrankungen beschrieben werden, wie Fibromyalgie, zeigen zusätzliche Symptome wie exzessive Ermüdung oder Mattigkeit (Miehlke und Späth 2003). Somit könnte angenommen werden, dass die Arbeitsleistung außer durch die Schmerzen auch durch die weiteren Symptome beeinträchtigt wird. Dieser Vermutung widersprechen jedoch teilweise die hohen Korrelationswerte von Schmerzintensität mit Funktionseinschränkungen, die Moore et al. 2010 bei Fibromyalgiepatienten nachweisen konnten.

Ergänzend wirken sich je nach Beruf Schmerzen und Funktionseinschränkungen unterschiedlich auf die jeweiligen Betroffenen aus. So mussten laut Breivik et al. 2006 16% der von chronischen Schmerzen betroffenen Patienten ihre Tätigkeitsschwerpunkte ändern.

Diese zahlreichen Faktoren, die bei arbeitsbezogenen Endpunkten eine Rolle spielen, machen sie zu einer schwer zu erfassenden und schwer zu interpretierenden Kategorie. Auch darin ist möglicherweise die mangelnde Präsenz dieses Untersuchungsfeldes begründet.

4.4 Erfassung von arbeitsbezogenen Endpunkten in zukünftigen Arbeiten

Für zukünftige Studien stellt sich die Frage, wie arbeitsbezogene Endpunkte untersucht und berichtet werden sollten. An erster Stelle sollte hierbei die Homogenisierung der Durchführung und Berichterstattung stehen. Daher sollte bei dem Studiendesign darauf geachtet werden, dass bereits validierte und bekannte Instrumente (Abschnitt 1.5) zum Einsatz kommen. Falls die Notwendigkeit gesehen würde, ein selbst entwickeltes Verfahren zu verwenden, so sollte dieses zusätzlich genutzt werden und nicht die validierten Instrumente ersetzen. Selbst entwickelte Instrumente sollten zuvor auf ihre Eignung hin untersucht und die zugehörigen Ergebnisse veröffentlicht werden.

In den 16 gefundenen, arbeitsbezogenen Endpunkten (Abschnitt 3.3) ließen sich 4 allgemeine Interessenbereiche erkennen:

- Probleme (Beeinträchtigung / Interferenz) bei der Arbeit
- Arbeitsproduktivität
- Arbeitsfehlzeiten
- Beschäftigungsstatus / Arbeitsfähigkeit

Die ersten beiden Themen sind sich sehr ähnlich. Es ist jedoch nicht möglich, diese in einer Meta-Analyse zu vereinen, da nicht bekannt ist, wie der genaue Zusammenhang in einem echten Feldversuch aussähe. So wären Unterschiede bei der Interpretation durch die Befragten, verursacht durch psychologische oder linguistische Variation des Verständnisses, denkbar. Probleme bei der Arbeit könnten eher mit dem Arbeitsverlauf dagegen die Arbeitsproduktivität eher mit dem Arbeitsergebnis assoziiert werden. Dabei ist ebenfalls zu berücksichtigen, dass Patienten das Gefühl haben könnten, trotz bestehender Behinderungen bei der Arbeit die gleiche Arbeitsproduktivität erreichen können wie ihre gesunden Kollegen. Als Argumentationshilfe könnten hier erhöhtes Eigenengagement oder eventuelle Hilfsmittel auf der Arbeit angeführt werden.

Der Bereich der Beeinträchtigung bei der Arbeit ist der im Moment am häufigsten und homogensten beschriebene Endpunkt. Zahlreiche Befragungsbögen machen Gebrauch von dieser Formulierung. Die meisten halten die Patienten dazu an, eine Wertung auf einer Punktskala von 0 bis 10 abzugeben, wobei 0 keine Probleme und 10 maximale Probleme angibt. Fragebögen welche diese Form der Befragung nutzen sind BPI, SDS, MAF, FIQ, WPS-RA. Die Unterpunkte dieser Instrumente sind gut miteinander zu vergleichen. Demgegenüber möchten die SF-36 und der DASH diesen Endpunk auf einer 0 bis 5 Skala bewertet sehen. Die Kombination von verschieden langen Skalen ist jedoch problematisch und nur über den Weg der Standardisierung mit seinen Limitationen zu bewerkstelligen. Wieder ist nicht klar, ob Patienten vorhersehbar auf unterschiedliche Skalen reagieren. Da die primären Endpunkte häufig nicht in den Bereich der Arbeitsmedizin fallen, ist es daher für die Autoren und auch für die Patienten einfacher einen Fragebogen zu verwenden, der bereits ihr zentrales Forschungsthema, zum Beispiel Schmerzen, mit abdeckt. Daher sind hier wegen ihrer Verbreitung das „Brief Pain Inventory“ oder das „Short Form 36 Health Survey Questionnaire“ zu empfehlen.

Die Produktivität am Arbeitsplatz wurde vornehmlich mit einer visuellen Analogskala (VAS) gemessen (Kavanaugh et al. 2006, Kavanaugh et al. 2013,

van der Heijde et al. 2006). Nur Skljarevski et al. 2010b verwendeten einen Unterpunkt des WPAI, welcher mit einem Punktesystem ähnlich dem BPI arbeitet. VAS können unterschiedlich gestaltet sein. In dieser Dissertation wurden VAS untersucht, welche eine Bewertung auf einer 0 bis 10 Skala ermöglichten. So konnten die Ergebnisse der verschiedener Studien leicht kombiniert werden. Zwei der drei Studien, die eine VAS über Arbeitsproduktivität untersuchten, berichteten ihre Ergebnisse als Median- und Interquartilangaben (Kavanaugh et al. 2006, van der Heijde et al. 2006). Dies lässt vermuten, dass die individuellen Ergebnisse der Patienten nicht normal verteilt waren, also die Grundvoraussetzung für eine Beschreibung der Daten mithilfe von Mittelwerten und Standardabweichungen nicht erfüllt waren (Higgins und Green 2011). Leider erklärten die Autoren ihr Vorgehen nicht, so dass eine Interpretation erschwert ist. Für die allgemeine Interpretation der Studien wäre es daher sinnvoll in zukünftigen Studien bereits bestehende Skalen zu verwenden und die Gründe der gewählten Präsentationsform zu erläutern.

In der Kategorie Arbeitsfehlzeiten fiel eine außerordentliche Heterogenität sowohl zwischen den beobachteten Zeiträumen als auch zwischen den Formen der Berichterstattung auf. In 13 Studien wurden Arbeitsfehlzeiten untersucht:

- Drei Studien berichteten keine Ergebnisse zu diesem Endpunkt (Licciardone et al. 2003, Barkham et al. 2010, Skljarevski et al. 2010a)
- Die Ergebnisse dreier weiterer Studien waren unbrauchbar, da aufgrund mangelnder Berichterstattung die Standardabweichungen nicht geschätzt werden konnten (Egsmose et al. 1997, Kavanaugh et al. 2006, Kavanaugh et al. 2009a).
- Drei verwertbare Studien berichteten im Rahmen des FIQ oder des WPAI über Fehlzeiten der vergangen Woche (Bennett et al. 2003 & 2005, Skljarevski et al. 2010a, Straube et al. 2011).

- Zwei Arbeiten untersuchten die Anzahl der Arbeitsfehltage während der Studie (Bejarano et al. 2008, van der Heijde et al. 2006).
- Smolen et al. 2006 berichteten Arbeitsfehltage kategorisch, indem aufgeführt wurde, wie viele Patienten jeweils 0, 0-10 und über 10 Arbeitstage im Verlauf der Studie verloren hatten.
- Albert et al. 2013 untersuchten Arbeitszeitverluste im vergangenen Jahr.

Aufgrund der heterogenen Berichterstattung der verschiedenen Studienautoren konnte keine dominante Präsentationsform für den Endpunkt „verlorene Arbeitszeit“ ausgemacht werden. Daher ist es wichtig, dass bei Veröffentlichung der Studienergebnisse auf genaue Beschreibung der Methodik und Ergebnisse geachtet wird. Es könnte weiterhin von Vorteil sein die Rohdaten für weitere Analysen zusätzlich zu veröffentlichen.

Beim Erfassen des Beschäftigungsstatus interessiert man sich für den tatsächlichen, derzeitigen Beschäftigungsgrad (z.B. Vollzeitbeschäftigung, Halbtagsbeschäftigung, keine Erwerbstätigkeit), während mit der Arbeitsfähigkeit zumeist eine Einschätzung, zu welchem Maß ein Patient theoretisch arbeiten könnte (z.B. voll arbeitsfähig, nicht arbeitsfähig), erfasst werden soll. In den betrachteten Studien wurde diese Art von Endpunkt ausschließlich kategorisch berichtet (Bejarano et al. 2008, Carlsson und Sjölund 2001, Kavanaugh et al. 2006, Lehmann et al. 1986, Manchikanti et al. 2010). Die einfachste Möglichkeit diesen Endpunkt zu untersuchen ist mithilfe einer binären Kategorisierung (arbeitsfähig und arbeitsunfähig), wie sie in dieser Arbeit gewählt wurde (Abschnitt 3.4.7). Der Beschäftigungsstatus sollte wegen der Aussagekraft einer möglichen patientenbezogenen Prognose möglichst detailliert beschrieben werden. So ist es bei diesen qualitativen Kategorien notwendig, eine Berichterstattung zu wählen, die die Entwicklung der einzelnen Patientengruppen nachvollziehbar macht. Dazu muss die Anzahl der sich zu Studienbeginn in der jeweiligen Kategorie befindlichen Probanden bekannt sein. Weiterhin ist eine genau Dokumentation über den Wechsel einzelner Probanden zwischen den unterschiedlichen Kategorien erforderlich.

4.5 Ansprechraten für arbeitsbezogene Endpunkte und die Berichterstattung des Behandlungseffektes

Wie in Abschnitt 3.5.2 gezeigt, gibt es einen starken Zusammenhang zwischen arbeitsbezogenen Endpunkten und schmerzbezogenen Endpunkten im Bereich Schmerzerkrankungen. Moore et al. 2010b argumentieren, dass bei Schmerzbehandlungen die durchschnittlichen Mittelwertsdifferenzen im klinischen Alltag nicht ausreichten, um Nutzen und Risikoabschätzungen für individuelle Behandlungen abzuwägen. Oft profitierte ein Teil der Patienten erheblich von einer Behandlung, während ein anderer Teil kaum Effekte berichten kann. Die gemeinsam gebildeten Mittelwerte gäben den Nutzen für die empfänglichen Patienten nicht wieder. Bereits in der Arbeit von McQuay et al. 1996 wird bestätigt, dass die absoluten Werte der Schmerzabnahme weder für Placebo noch für aktive Behandlungen normalverteilt seien und daher eine Darstellung des Effektes als Normalverteilungsderivat mit Mittelwertsdifferenz und Standardabweichung verzerrend sei. Dies sei mit der Transformierung der Daten in ein binäres System mithilfe von definierten Schwellenwerten vermeidbar. Als zu verwendende Schwellen definierte die IMMPACT-Arbeitsgruppe (Dworkin et al. 2009) für chronische Schmerzen eine 30%ige Verbesserung als ein moderates beziehungsweise eine 50%ige Verbesserung als ein bedeutendes Ansprechen auf die Behandlung. O'Brien et al. stimmen dem 2010 in ihrer Studie zu. Sie fanden heraus, dass sich die Patienten von ihrer Schmerzbehandlung eine Verbesserung von 50 bis 60 Prozent erhofften.

Straube et al. untersuchten 2010 in ihrer Meta-Analyse arbeitsbezogene Endpunkte mithilfe von Ansprechraten. Dabei fiel auf, dass einfach definierte Endpunkte wie Schmerz und Schlaf gut mit Ansprechraten zu untersuchen waren, komplexere Definitionen wie FIQ-Unterpunkte oder Angst betreffende Befragungen jedoch mangelnde Sensitivität aufwiesen.

Eine interessante Fragestellung für zukünftige Arbeiten ist es, wie arbeitsbezogene Endpunkte und ihre Behandlungseffekte verteilt sind und wie diese nun am besten mathematisch zu modellieren seien. Sind diese nicht normalverteilt, so könnte es sinnvoll sein, Ansprechraten, wie sie vergleichbar für

Schmerzendpunkte zur Anwendung kommen, auch für arbeitsbezogene Endpunkte zu definieren.

5. Zusammenfassung

Um einen Überblick über die Datenlage bezogen auf arbeitsrelevante Endpunkte in randomisierten, klinischen Studien zur Behandlung chronischer Schmerzerkrankungen und schmerzhafter rheumatischer Erkrankungen zu gewinnen, wurde ein systematisches Review erstellt.

Die systematische Literaturrecherche erfolgte mithilfe der Datenbank der „US National Library of Medicine“ (MEDLINE®/PubMED®). Die Ergebnisse der geeigneten Studien wurden mithilfe von Meta-Analysen untersucht.

31 Arbeiten mit insgesamt 11434 Patienten enthielten Informationen über arbeitsbezogene Endpunkte.

Diese Publikationen machen 0,23% aller Veröffentlichungen aus, die sich mit placebokontrollierten Behandlungsstudien zu chronischen Schmerzerkrankungen und rheumatischen Erkrankungen beschäftigten. Obwohl in diesem Bereich die Veröffentlichungen pro Jahr exponentiell zunahmen, blieb der relative Anteil der Studien, die arbeitsbezogene Endpunkte berichteten, im letzten Jahrzehnt annähernd gleich.

Bei einer umfassenden Meta-Analyse aller arbeitsbezogenen Endpunkte, bei der die Ergebnisse der einzelnen Studien standardisiert werden mussten, konnte ein allgemeiner Behandlungserfolg von 0,35, ausgedrückt als standardisierte Mittelwertsdifferenz (SMD), gefunden werden (95%-KI: 0,21; 0,50).

Weitere Meta-Analysen konnten für die Untersuchung von Beeinträchtigungen bei der Arbeit eine Verbesserung von 0,62 Punkten auf einer Skala von 0-10 nachweisen (95%-KI: 0,45; 0,79).

Im Bereich der Arbeitsproduktivität konnte eine Verbesserung von 0,53, ausgedrückt als SMD, festgestellt werden (95%-KI: 0,18; 0,88). Eine Analyse nur solcher Studien, welche Arbeitsproduktivität mit einer visuellen Analogskala

untersuchten, ergab eine Verbesserung von 2,13 auf einer Skala von 0 bis 10 (95%-KI: 1,69; 2,58).

Drei Studien über die Behandlung von chronischen Rückenschmerzen mit Duloxetin berichteten unter Anwendung des „Brief Pain Inventory“ über die Teilfrage „Beeinträchtigung der normalen Arbeit“. Die gepoolte Verbesserung betrug 0,55 auf einer Skala von 0 bis 10 (95%-KI: 0,25; 0,86).

Die Endpunkte, die Arbeitsfehlzeiten erfassten, waren heterogen und die Ergebnisse der Studien stellten sich als uneinheitlich dar.

Ebenso konnten Analysen, das Beschäftigungsverhältnis betreffend, zu keinen eindeutigen Ergebnissen kommen.

Bei Regressionen von standardisierten, schmerzbezogenen mit standardisierten, arbeitsbezogenen Endpunkten fand sich ein statistisch signifikanter linearer Zusammenhang. Weiterhin wurde die Beziehung zwischen Schmerzendpunkten und arbeitsbezogenen Endpunkten untersucht. Eine Meta-Analyse mit standardisierten Schmerzendpunkten ergab eine allgemeine Verbesserung von schmerzbezogenen Endpunkten von 0,40, ausgedrückt als SMD (95%-KI: 0,25; 0,55).

Im Bereich von chronischen Schmerzen und rheumatischen Erkrankungen werden in den letzten Jahren häufig Ansprechraten, bezogen auf die Behandlungen und jeweiligen Endpunkte, erfasst. Die erstellten Meta-Analysen der schmerzbezogenen Ansprechraten sowie die der Ansprechraten auf die Kriterien des „American College of Rheumatology“ (ACR) wiesen ausnahmslos signifikante Verbesserungen in den Behandlungsgruppen auf. So konnten die schmerzbezogenen 30%- und 50%-Ansprechraten jeweils eine 33%-ige beziehungsweise 46%-ige Überlegenheit gegenüber Placebo vorweisen. Für die Ansprechraten nach den Kriterien des American College of Rheumatology konnten für die 20%-, 50%- und 70%-Schwellen deutliche Verbesserungen um 184%, 237% und 232% gegenüber Placebo identifiziert werden.

Regressionen der schmerzbezogenen Ansprechraten mit den standardisierten Mittelwertsdifferenzen der arbeitsbezogenen Endpunkte konnten ebenfalls einen statistisch signifikanten Zusammenhang nachweisen.

Für zukünftige Arbeiten stellt sich die Frage, wie die Ergebnisse von arbeitsbezogenen Endpunkten verteilt sind und wie diese daraus folgend sinnvoll berichtet werden sollten. Folgen sie einer Normalverteilung, so ist die klassische Berichterstattung mit Mittelwert und Standardabweichung angezeigt. Folgen sie jedoch - wie viele Schmerzendpunkte - einer anderen Verteilung, so könnten auch hier Ansprechraten die geeignete Präsentationsform sein.

Literaturverzeichnis

Affaitati G, Fabrizio A, Savini A, Lerza R, Tafuri E, Costantini R, Lapenna D, Giamberardino MA (2009): A randomized, controlled study comparing a lidocaine patch, a placebo patch, and anesthetic injection for treatment of trigger points in patients with myofascial pain syndrome: evaluation of pain and somatic pain thresholds. Clin Ther 31, 705-720

Albert HB, Sorensen JS, Christensen BS, Manniche C (2013): Antibiotic treatment in patients with chronic low back pain and vertebral bone edema (Modic type 1 changes): a double-blind randomized clinical controlled trial of efficacy. Eur Spine J 22, 697-707

Al-Smadi J, Warke K, Wilson I, Cramp AF, Noble G, Walsh DM, Lowe-Strong AS (2003): A pilot investigation of the hypoalgesic effects of transcutaneous electrical nerve stimulation upon low back pain in people with multiple sclerosis. Clin Rehabil 17, 742-749

Altman DG: Practical statistics for medical research. Chapman & Hall, London 1991

Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, Christy W, Cooke TD, Greenwald R, Hochberg M et al. (1986): Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum 29, 1039-1049

Antal A, Terney D, Kühnl S, Paulus W (2010): Anodal transcranial direct current stimulation of the motor cortex ameliorates chronic pain and reduces short intracortical inhibition. J Pain Symptom Manage 39, 890-903

Antoni C, Krueger GG, deVlam K, Birbara C, Beutler A, Guzzo C, Zhou B, Dooley LT, Kavanaugh A (2005): Infliximab improves signs and symptoms of psoriatic arthritis: Results of the IMPACT 2 trial. Ann Rheum Dis 64, 1150-1157.

Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS et al. (1988): The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 31, 315-324

Arnold LM, Rosen A, Pritchett YL, D'Souza DN, Goldstein DJ, Iyengar S, Wernicke JF (2005): A randomized, double-blind, placebo-controlled trial of duloxetine in the treatment of women with fibromyalgia with or without major depressive disorder. Pain 119, 5-15

Arnold LM, Russell IJ, Diri EW, Duan WR, Young JP Jr, Sharma U, Martin SA, Barrett JA, Haig G (2008): A 14-week, randomized, double-blinded, placebo-controlled monotherapy trial of pregabalin in patients with fibromyalgia. J Pain 2008 9, 792-805.

Arnold LM, Clauw DJ, Wohlreich MM, Wang F, Ahl J, Gaynor PJ, Chappell AS (2009): Efficacy of duloxetine in patients with fibromyalgia: pooled analysis of 4 placebo-controlled clinical trials. Prim Care Companion J Clin Psychiatry 11, 237-244

Arnold LM, Zlateva G, Sadosky A, Emir B, Whalen E (2011): Correlations between fibromyalgia symptom and function domains and patient global impression of change: a pooled analysis of three randomized, placebo-controlled trials of pregabalin. Pain Med 12, 260-267

Bagust A, Boland A, Hockenhull J, Fleeman N, Greenhalgh J, Dundar Y, Proudlove C, Kennedy T, Moots R, Williamson P, Dickson R (2009): Rituximab for the treatment of rheumatoid arthritis. Health Technol Assess 13 Suppl 2, 23-29

Baranowski AP, Lee J, Price C, Hughes J (2014): Pelvic pain: a pathway for care developed for both men and women by the British Pain Society. Br J Anaesth 112, 452-459

Barkham N, Coates LC, Keen H, Hensor E, Fraser A, Redmond A, Cawkwell L, Emery P (2010): Double-blind placebo-controlled trial of etanercept in the prevention of work disability in ankylosing spondylitis. Ann Rheum Dis 69, 1926-1928

Baron R, Freynhagen R, Tölle TR, Cloutier C, Leon T, Murphy TK, Phillips K, A0081007 Investigators (2010): The efficacy and safety of pregabalin in the treatment of neuropathic pain associated with chronic lumbosacral radiculopathy. Pain 150, 420-427

Bejarano V, Quinn M, Conaghan PG, Reece R, Keenan AM, Walker D, Gough A, Green M, McGonagle D, Adebajo A et al.(2008): Effect of the early use of the anti-tumor necrosis factor adalimumab on the prevention of job loss in patients with early rheumatoid arthritis. Arthritis Rheum 59, 1467-1474

Bellato E, Marini E, Castoldi F, Barbasetti N, Mattei L, Bonasia DE, Blonna D (2012): Fibromyalgia syndrome: etiology, pathogenesis, diagnosis, and treatment. Pain Res Treat 2012, 426130

Bennell K, Coburn S, Wee E, Green S, Harris A, Forbes A, Buchbinder R (2007): Efficacy and cost-effectiveness of a physiotherapy program for chronic rotator cuff pathology: a protocol for a randomised, double-blind, placebo-controlled trial. BMC Musculoskelet Disord 8, 86

Bennett RM, Kamin M, Karim R, Rosenthal N (2003): Tramadol and acetaminophen combination tablets in the treatment of fibromyalgia pain: a double-blind, randomized, placebo-controlled study. Am J Med 114, 537– 545.

Bennett RM, Schein J, Kosinski MR, Hewitt DJ, Jordan DM, Rosenthal NR (2005): Impact of fibromyalgia pain on health-related quality of life before and after treatment with tramadol/acetaminophen. Arthritis Rheum. 53, 519-527

Beurskens AJ, de Vet HC, Köke AJ, Regtop W, van der Heijden GJ, Lindeman E, Knipschild PG (1997): Efficacy of traction for nonspecific low back pain. 12-week and 6-month results of a randomized clinical trial. Spine (Phila Pa 1976) 22, 2756-2762

Bjordal JM, Lopes-Martins RA, Joensen J, Couppe C, Ljunggren AE, Stergioulas A, Johnson MI (2008): A systematic review with procedural assessments and meta-analysis of low level laser therapy in lateral elbow tendinopathy (tennis elbow). BMC Musculoskelet Disord 9, 75

Bombardier C, Raboud J (1991): A comparison of health-related quality-of-life measures for rheumatoid arthritis research. The Auranofin Cooperating Group. Control Clin Trials 12, 243S-256S

Borg G, Allander E, Berg E, Brodin U, From A, Trang L (1991): Auranofin treatment in early rheumatoid arthritis may postpone early retirement. Results from a 2-year double blind trial. J Rheumatol 18, 1015-1020

Bradley LA, Bennett R, Russell IJ, Wohlreich MM, Chappell AS, Wang F, D'Souza DN, Moldofsky H (2010): Effect of duloxetine in patients with fibromyalgia: tiredness subgroups. Arthritis Res Ther 12, R141

Braun J, Brandt J, Listing J, Zink A, Alten R, Golder W, Gromnica-Ihle E, Kellner H, Krause A, Schneider M, Sörensen H, Zeidler H, Thriene W, Sieper J (2002): Treatment of active ankylosing spondylitis with infliximab: a randomised controlled multicentre trial. Lancet 359, 1187-1193

Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, Westlake L (1992): Validating the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ 305, 160-164

Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D (2006): Survey of chronic pain in Europe: Prevalence, impact on daily life, and treatment. Eur J Pain 10, 287-333

Bresnihan B (2002): Anakinra as a new therapeutic option in rheumatoid arthritis: clinical results and perspectives. Clin Exp Rheumatol 20, S32-S34

Brown BB (1968): Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts. RAND Corporation P, 3925

Brox JI, Storheim K, Grotle M, Tveito TH, Indahl A, Eriksen HR (2008): Systematic review of back schools, brief education, and fear-avoidance training for chronic low back pain. Spine J 8, 948-958

Bruyere O, Delferriere D, Roux C, Wark JD, Spector T, Devogelaer JP, Brixen K, Adami S, Fechtenbaum J, Kolta S, Reginster JY (2008): Effects of strontium ranelate on spinal osteoarthritis progression. Ann Rheum Dis 67, 335-339

Bunzli S, Gillham D, Esterman A (2011): Physiotherapy-provided operant conditioning in the management of low back pain disability: A systematic review. Physiother Res Int 16, 4-19

Busanich BM, Verscheure SD (2006): Does McKenzie therapy improve outcomes for back pain? J Athl Train 41, 117-119

Busch AJ, Barber KA, Overend TJ, Peloso PMJ, Schachter CL. (2007) Exercise for treating fibromyalgia syndrome. Cochrane Database of Systematic Reviews 4, CD003786

Cade R, Stein G, Pickering M, Schlein E, Spooner G (1976): Low dose, long-term treatment of rheumatoid arthritis with azathioprine. South Med J 69, 388-392

Cardiel MH, Tak PP, Bensen W, Burch FX, Forejtova S, Badurski JE, Kakkar T, Bevirt T, Ni L, McCroskery E, Jahreis A, Zack DJ (2010): A phase 2 randomized, double-blind study of AMG 108, a fully human monoclonal antibody to IL-1R, in patients with rheumatoid arthritis. Arthritis Res Ther 12, R192

Carlier BE, Schuring M, Lötters FJ, Bakker B, Borgers N, Burdorf A (2013): The influence of re-employment on quality of life and self-rated health, a longitudinal study among unemployed persons in the Netherlands. BMC Public Health 13, 503

Carlsson CP, Sjölund BH (2001): Acupuncture for chronic low back pain: a randomized placebo-controlled study with long-term follow-up. Clin J Pain 17, 296-305

Chaparro LE, Furlan AD, Deshpande A, Mailis-Gagnon A, Atlas S, Turk DC (2013): Opioids compared to placebo or other treatments for chronic low-back pain. Cochrane Database of Systematic Reviews 8, CD004959

Chappell AS, Bradley LA, Wiltse C, Detke MJ, D'Souza DN, Spaeth M (2008): A six-month double-blind, placebo-controlled, randomized clinical trial of duloxetine for the treatment of fibromyalgia. Int J Gen Med 1, 91-102

Chappell AS, Ossanna MJ, Liu-Seifert H, Iyengar S, Skljarevski V, Li LC, Bennett RM, Collins H (2009): Duloxetine, a centrally acting analgesic, in the treatment of patients with osteoarthritis knee pain: a 13-week, randomized, placebo-controlled trial. Pain 146, 253-260

Chappell AS, Desaiah D, Liu-Seifert H, Zhang S, Skljarevski V, Belenkov Y, Brown JP (2011): A double-blind, randomized, placebo-controlled study of the efficacy and safety of duloxetine for the treatment of chronic pain due to osteoarthritis of the knee. Pain Pract 11, 33-41

Chen J, Veras MM, Liu C, Lin J (2013): Methotrexate for ankylosing spondylitis. Cochrane Database Syst Rev 2, CD004524

Chen YF, Jobanputra P, Barton P, Jowett S, Bryan S, Clark W, Fry-Smith A, Burls A (2006): A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness. Health Technol Assess 10, 42

Chiu CK, Low TH, Tey YS, Singh VA, Shong HK (2011): The efficacy and safety of intramuscular injections of methylcobalamin in patients with chronic nonspecific low back pain: a randomised controlled trial. Singapore Med J 52, 868-873

Chinn S (2000): A simple method for converting an odds ratio to effect size for use in meta-analysis. Stat Med 19, 3127-3131

Chou R, Huffman LH; American Pain Society; American College of Physicians (2007): Nonpharmacologic therapies for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. Ann Intern Med 147, 492-504

Clare HA, Adams R, Maher CG (2004): A systematic review of efficacy of McKenzie therapy for spinal pain. Aust J Physiother 50, 209-216

Clark W, Jobanputra P, Barton P, Burls A (2004): The clinical and cost-effectiveness of anakinra for the treatment of rheumatoid arthritis in adults: a systematic review and economic analysis. Health Technol Assess 8, 18

Clopper CJ, Pearson ES (1934): The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26, 404–416

Coghlan JA, Buchbinder R, Green S, Johnston RV, Bell SN (2008): Surgery for rotator cuff disease. Cochrane Database Syst Rev 1, CD005619

Cohen J: Statistical Power Analysis for the Behavioral Sciences, 2. Auflage; Lawrence Erlbaum Associates, Hillsdale, New Jersey 1988

Cole JC, Li T, Lin P, MacLean R, Wallenstein GV (2008): Treatment impact on estimated medical expenditure and job loss likelihood in rheumatoid arthritis: re-examining quality of life outcomes from a randomized placebo-controlled clinical trial with abatacept. Rheumatology 47, 1044-1050

Crofford LJ, Rowbotham MC, Mease PJ, Russell IJ, Dworkin RH, Corbin AE, Young JP Jr, LaMoreaux LK, Martin SA, Sharma U; Pregabalin 1008-105 Study Group (2005): Pregabalin for the treatment of fibromyalgia syndrome: results of a randomized, double-blind, placebo-controlled trial. Arthritis Rheum 52, 1264-1273

Cummings P (2011): Arguments for and Against Standardized Mean Differences (Effect Sizes). Arch Pediatr Adolesc Med 165, 592-596

Davies GM, Watson DJ, Bellamy N (1999): Comparison of the responsiveness and relative effect size of the western Ontario and McMaster Universities Osteoarthritis Index and the short-form Medical Outcomes Study Survey in a randomized, clinical trial of osteoarthritis patients. Arthritis Care Res 12, 172-179

Deare JC, Zheng Z, Xue CCL, Liu JP, Shang J, Scott SW, Littlejohn G (2013): Acupuncture for treating fibromyalgia. Cochrane Database Syst Rev 5, CD007070

Debono DJ, Hoeksema LJ, Hobbs RD (2013): Caring for patients with chronic pain: pearls and pitfalls. J Am Osteopath Assoc 113, 620-627

Deeks JJ (2002): Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. Stat Med 21, 1575-1600

Deeks JJ, Higgins JPT (2010): Statistical algorithms in Review Manager 5
on behalf of the Statistical Methods Group. The Cochrane Collaboration, Available from www.cochrane-handbook.org

Dworkin RH, Turk DC, McDermott MP, Peirce-Sandner S, Burke LB, Cowan P, Farrar JT, Hertz S, Raja SN, Rappaport BA, Rauschkolb C, Sampaio C (2009): Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations Pain 146, 238-244

Egsmose C, Hansen TM, Andersen LS, Beier JM, Christensen L, Ejstrup L, Peters ND, van der Heijde DM (1997): Limited effect of sulphasalazine treatment in reactive arthritis. A randomised double blind placebo controlled trial. Ann Rheum Dis 56, 32-36

Falco FJ, Manchikanti L, Datta S, Wargo BW, Geffert S, Bryce DA, Atluri S, Singh V, Benyamin RM, Sehgal N, Ward SP, Helm S 2nd, Gupta S, Boswell MV (2012): Systematic review of the therapeutic effectiveness of cervical facet joint interventions: an update. Pain Physician 15, E839-E868

Fraser AD, van Kuijk AW, Westhovens R, Karim Z, Wakefield R, Gerards AH, Landewé R, Steinfeld SD, Emery P, Dijkmans BA, Veale DJ (2005): A randomised, double blind, placebo controlled, multicentre trial of combination therapy with methotrexate plus ciclosporin in patients with active psoriatic arthritis. Ann Rheum Dis 64, 859-864

Freeman BJ (2006): IDET: a critical appraisal of the evidence. Eur Spine J 15, S448-S457

Furlan AD, Imamura M, Dryden T, Irvin E (2008): Massage for low-back pain. Cochrane Database Syst Rev 4, CD001929

Furlan AD, Yazdi F, Tsertsvadze A, Gross A, Van Tulder M, Santaguida L, Cherkin D, Gagnier J, Ammendolia C, Ansari MT, Ostermann T, Dryden T, Doucette S, Skidmore B, Daniel R, Tsouros S, Weeks L, Galipeau J (2010): Complementary and alternative therapies for back pain II. Evid Rep Technol Assess 194, 1-764

Geisser ME, Wiggert EA, Haig AJ, Colwell MO (2005): A randomized, controlled trial of manual therapy and specific adjuvant exercise for chronic low back pain. Clin J Pain. 21, 463-470

Gibson JN, Waddell G (2005): Surgery for degenerative lumbar spondylosis. Cochrane Database Syst Rev 4, CD001352

Gilworth G, Emery P, Barkham N, Smyth MG, Helliwell P, Tennant A (2009): Reducing work disability in Ankylosing Spondylitis: development of a work instability scale for AS. BMC Musculoskelet Disord 10, 68

Goldenberg D, Mayskiy M, Mossey C, Ruthazer R, Schmid C (1996): A randomized, double-blind crossover trial of fluoxetine and amitriptyline in the treatment of fibromyalgia. Arthritis Rheum 39, 1852-1859

Han C, Smolen J, Kavanaugh A, St Clair EW, Baker D, Bala M (2008): Comparison of employability outcomes among patients with early or long-standing rheumatoid arthritis. Arthritis Rheum 59, 510-514

Hardy RJ, Thompson SG (1998): Detecting and describing heterogeneity in meta-analysis. Stat Med 17, 841-856.

Häuser W, Urrútia G, Tort S, Üçeyler N, Walitt B (2013): Serotonin and noradrenaline reuptake inhibitors (SNRIs) for fibromyalgia syndrome. Cochrane Database Syst Rev 1, CD010292

Hayden JA, van Tulder MW, Malmivaara A, Koes BW (2005): Exercise therapy for treatment of non-specific low back pain. Cochrane Database Syst Rev 3, CD000335

He D, Høstmark AT, Veiersted KB, Medbø JI (2005): Effect of intensive acupuncture on pain-related social and psychological variables for women with chronic neck and shoulder pain--an RCT with six month and three year follow up. Acupunct Med 23, 52-61

Hedges LV (1981): Distribution Theory for Glass's Estimator of Effect size and Related Estimators. Journal of Educational and Behavioral Statistics 6, 107-128

Hennigan S, Ackermann C, Kavanaugh A (2008): Adalimumab in ankylosing spondylitis: an evidence-based review of its place in therapy. Core Evid 2, 295-305

Hettenkofer HJ, Neudorf U, Hammer M: Entzündliche rheumatische Erkrankungen. In: Rheumatologie. 5. Auflage; hrsg. v. Hettenkofer HJ; Thieme Verlag, Stuttgart 2003, 59-159

Hewitson PJ, Debroe S, McBride A, Milne R (2000): Leflunomide and rheumatoid arthritis: a systematic review of effectiveness, safety and cost implications. J Clin Pharm Ther 25, 295-302

Heymans MW, van Tulder MW, Esmail R, Bombardier C, Koes BW (2004): Back schools for non-specific low-back pain. Cochrane Database Syst Rev 4, CD000261

Higgins JPT, Green S (editors) (2011): Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, Available from www.cochrane-handbook.org.

Higgins JPT, Thompson SG, Deeks JJ, Altman DG (2003): Measuring inconsistency in meta-analyses. BMJ 327, 557-560

Hill HF, Hill AG (1975): Ankylosing spondylitis: open long-term and double-blind crossover studies with naproxen. J Clin Pharmacol 15, 355-362

Hollis S, Campbell F (1999): What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ 319, 670-674

Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K (2009): Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database Syst Rev. 1, MR000006

Hudson JI, Arnold LM, Bradley LA, Choy EH, Mease PJ, Wang F, Ahl J, Wohlreich MM (2009): What makes patients with fibromyalgia feel better? Correlations between Patient Global Impression of Improvement and changes in clinical symptoms and function: a pooled analysis of 4 randomized placebo-controlled trials of duloxetine. J Rheumatol 36, 2517-2522

Hunter AM, Leuchter AF, Cook IA, Abrams M, Siegman BE, Furst DE, Chappell AS (2009): Brain functional changes and duloxetine treatment response in fibromyalgia: a pilot study. Pain Med 10, 730-738

Hutchinson AJ, Ball S, Andrews JC, Jones GG (2012): The effectiveness of acupuncture in treating chronic non-specific low back pain: a systematic review of the literature. J Orthop Surg Res 7, 36

Itoh K, Kitakoji H (2007): Acupuncture for chronic pain in Japan: a review. Evid Based Complement Alternat Med 4, 431-438

Iversen T, Solberg TK, Romner B, Wilsgaard T, Twisk J, Anke A, Nygaard O, Hasvold T, Ingebrigtsen T (2011): Effect of caudal epidural steroid or saline injection in chronic lumbar radiculopathy: multicentre, blinded, randomised controlled trial. BMJ 343, d5278

Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ (1996): Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 17, 1-12

Jarzem PF, Harvey EJ, Arcaro N, Kaczorowski J (2005): Transcutaneous Electrical Nerve Stimulation [TENS] for Chronic Low Back Pain. J Musculoskelet Pain 13, 3-9

Kavanaugh A, Antoni C, Mease P, Gladman D, Yan S, Bala M, Zhou B, Dooley LT, Beutler A, Guzzo C, Krueger GG (2006) Effect of infliximab therapy on employment, time lost from work, and productivity in patients with psoriatic arthritis. J Rheumatol 33, 2254-2259

Kavanaugh A, Smolen JS, Emery P, Purcaru O, Keystone E, Richard L, Strand V, van Vollenhoven RF (2009a): Effect of certolizumab pegol with methotrexate on home and work place productivity and social activities in patients with active rheumatoid arthritis. Arthritis Rheum 61, 1592-1600

Kavanaugh A, McInner I, Mease P, Krueger GG, Gladman D, Gomez-Reino J, Papp K, Zrubek J, Mudivarthy S, Mack M, Visvanathan S, Beutler A (2009b): Golimumab, a new human tumor necrosis factor alpha antibody, administered every four weeks as a subcutaneous injection in psoriatic arthritis: Twenty-four-week efficacy and safety results of a randomized, placebo-controlled study. Arthritis Rheum 60, 976-986

Kavanaugh A, McInnes IB, Krueger GG, Gladman D, Beutler A, Gathany T, Mack M, Tandon N, Han C, Mease P (2013): Patient-Reported Outcomes and the Association With Clinical Response in Patients With Active Psoriatic Arthritis Treated With Golimumab: Findings Through 2 Years of a Phase III, Multicenter, Randomized, Double-Blind, Placebo-Controlled Trial. Arthritis Care Res 65, 1666-1673

Kay J, Rahman MU (2010): Golimumab: A novel human anti-TNF-alpha monoclonal antibody for the treatment of rheumatoid arthritis, ankylosing spondylitis, and psoriatic arthritis. Core Evid 4, 159-170

Keystone E, van der Heijde D, Mason D Jr, Landewe R, van Vollenhoven R, Combe B, Emery P, Strand V, Mease P, Desai C, Pavelka K (2008): Certolizumab pegol plus methotrexate is significantly more effective than placebo plus methotrexate in active rheumatoid arthritis: findings of a fifty-two–week, phase III, multicenter, randomized, double-blind, placebo-controlled, parallel-group study. Arthritis Rheum 58, 3319–3329

Khadilkar A, Odebiyi DO, Brosseau L, Wells GA (2008): Transcutaneous electrical nerve stimulation (TENS) versus placebo for chronic low-back pain. Cochrane Database Syst Rev 4, CD003008

Kimel M, Revicki D, Rao S, Fryback D, Feeny D, Harnam N, Thompson C, Cifaldi M (2011): Norms-based assessment of patient-reported outcomes associated with adalimumab monotherapy in patients with ankylosing spondylitis. Clin Exp Rheumatol 29, 624-632

Kirveskari P, Alanen P (1984): Effect of occlusal treatment on sick leaves in TMJ dysfunction patients with head and neck symptoms. Community Dent Oral Epidemiol 12, 78-81

Kirwan JR, Bijlsma JWJ, Boers M, Shea B (2007): Effects of glucocorticoids on radiological progression in rheumatoid arthritis. Cochrane Database Syst Rev 1, CD006356

Kuijpers T, van Middelkoop M, Rubinstein SM, Ostelo R, Verhagen A, Koes BW, van Tulder MW (2011): A systematic review on the effectiveness of pharmacological interventions for chronic non-specific low-back pain. Eur Spine J 20, 40-50

Lane P (2008): Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. Pharm Stat 7, 193-106

Lehmann TR, Russell DW, Spratt KF, Colby H, Liu YK, Fairchild ML, Christensen S (1986): Efficacy of electroacupuncture and TENS in the rehabilitation of chronic low back pain patients. Pain 26, 277-290

Lerner D, Chang H, Rogers WH, Benson C, Chow W, Kim MS, Biondi D (2012): Imputing at-work productivity loss using results of a randomized controlled trial comparing tapentadol extended release and oxycodone controlled release for osteoarthritis pain. J Occup Environ Med 54, 933-938

Lewis S, Clarke M (2001): Forest plots: trying to see the wood and the trees. BMJ 322, 1479-1480

Licciardone JC, Stoll ST, Fulda KG, Russo DP, Siu J, Winn W, Swift J Jr (2003): Osteopathic manipulative treatment for chronic low back pain: a randomized controlled trial. Spine (Phila Pa 1976) 28, 1355-1362

Lindberg DAB (2000): Internet Access to the National Library of Medicine. Eff Clin Pract 4, 256-260

Listing J, Brandt J, Rudwaleit M, Zink A, Sieper J, Braun J (2004): Impact of anti-tumour necrosis factor alpha treatment on admissions to hospital and days of sick leave in patients with ankylosing spondylitis. Ann Rheum Dis 63, 1670-1672

Malottki K, Barton P, Tsourapas A, Uthman AO, Liu Z, Routh K, Connock M, Jobanputra P, Moore D, Fry-Smith A, Chen YF (2011): Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a tumour necrosis factor inhibitor: a systematic review and economic evaluation. Health Technol Assess 15, 14

Manchikanti L, Singh V, Falco FJ, Cash KA, Fellows B (2010): Comparative outcomes of a 2-year follow-up of cervical medial branch blocks in management of chronic neck pain: a randomized, double-blind controlled trial. Pain Physician 13, 437-450

Manek NJ, MacGregor AJ (2005): Epidemiology of back disorders: prevalence, risk factors, and prognosis. Curr Opin Rheumatol 17, 134-140

Manheimer E, White A, Berman B, Forys K, Ernst E (2005): Meta-analysis: acupuncture for low back pain. Ann Intern Med 142, 651-663

Maniadakis N, Gray A (2000): The economic burden of back pain in the UK. Pain 84, 95-103

Markenson JA, Croft J, Zhang PG, Richards P (2005): Treatment of persistent pain associated with osteoarthritis with controlled-release oxycodone tablets in a randomized controlled clinical trial. Clin J Pain 21, 524-535

McLeod C, Bagust A, Boland A, Dagenais P, Dickson R, Dundar Y, Hill RA, Jones A, Mujica Mota R, Walley T (2007): Adalimumab, etanercept and infliximab for the treatment of ankylosing spondylitis: a systematic review and economic evaluation. Health Technol Assess 11, 28

McQuay H, Carroll D, Moore A (1996): Variation in the placebo effect in randomised controlled trials of analgesics: all is as blind as it seems. Pain 64, 331-335

Mease PJ, Russell IJ, Arnold LM, Florian H, Young JP Jr, Martin SA, Sharma U (2008): A randomized, double-blind, placebo-controlled, phase III trial of pregabalin in the treatment of patients with fibromyalgia. J Rheumatol 35, 502-514

Meireles SM, Jones A, Jennings F, Suda AL, Parizotto NA, Natour J (2010): Assessment of the effectiveness of low-level laser therapy on the hands of patients with rheumatoid arthritis: a randomized double-blind controlled trial. Clin Rheumatol 29, 501-509

Miehlke RK, Schmidt K: Degenerative rheumatische Erkrankungen. In: Rheumatologie. 5. Auflage; hrsg. v. Hettenkofer HJ; Thieme Verlag, Stuttgart 2003, 160-184

Miehlke RK, Späth M: Extraartikulärer Rheumatismus. In: Rheumatologie 5. Auflage; hrsg. v. Hettenkofer HJ; Thieme Verlag, Stuttgart 2003, 185-231

Mittendorf T, Dietz B, Sterz R, Cifaldi MA, Kupper H, von der Schulenburg JM (2008): Personal and economic burden of late-stage rheumatoid arthritis among patients treated with adalimumab: an evaluation from a patient's perspective. Rheumatology (Oxford) 47, 188-193

Moore RA, Straube S, Wiffen PJ, Derry S, McQuay HJ (2009): Pregabalin for acute and chronic pain in adults. Cochrane Database Syst Rev 3, CD007076

Moore RA, Straube S, Paine J, Phillips CJ, Derry S, McQuay HJ (2010a): Fibromyalgia: Moderate and substantial pain intensity reduction predicts improvement in other outcomes and substantial quality of life gain. Pain 149, 360-364

Moore RA, Eccleston C, Derry S, Wiffen P, Bell RF, Straube S, McQuay H; ACTINPAIN Writing Group of the IASP Special Interest Group on Systematic Reviews in Pain Relief; Cochrane Pain, Palliative and Supportive Care Systematic Review Group Editor (2010b): "Evidence" in chronic pain--establishing best practice in the reporting of systematic reviews. Pain 150, 386-389

Moore RA, Derry S, Aldington D, Cole P, Wiffen PJ (2012a): Amitriptyline for neuropathic pain and fibromyalgia in adults. Cochrane Database Syst Rev 12, CD008242

Moore RA, Straube S, Eccleston C, Derry S, Aldington D, Wiffen P, Bell RF, Hamunen K, Phillips C, McQuay H (2012b): Estimate at your peril: imputation methods for patient withdrawal can bias efficacy outcomes in chronic pain trials using responder analyses. Pain 153, 265-268

Moreland LW (2004): Drugs that block tumour necrosis factor: experience in patients with rheumatoid arthritis. Pharmacoeconomics. 22(2 Suppl 1), 39-53

Muazzam AG, Mansoor A, Ali L, Siddiqi S, Hameed A, Ajmal M, Mazhar K (2013): Association of HLA-DRB1 and -DQB1alleles and haplotypes with rheumatoid arthritis in a Pakistani population. Arthritis Res Ther 15, R95

Navarro-Millán I, Singh JA, Curtis JR (2012): Systematic review of tocilizumab for rheumatoid arthritis: a new biologic agent targeting the interleukin-6 receptor. Clin Ther 34, 788-802

Negi S, Juyal G, Senapati S, Prasad P, Gupta A, Singh S, Kashyap S, Kumar A, Kumar U, Gupta R et al. (2013): A genome-wide association study reveals ARL15, a novel non-HLA susceptibility gene for Rheumatoid arthritis in north Indians. Arthritis Rheum 65, 3026-2035

Nepom GT, Byers P, Seyfried C, Healey LA, Wilske KR, Stage D, Nepom BS (1989): HLA genes associated with rheumatoid arthritis. Identification of susceptibility alleles using specific oligonucleotide probes. Arthritis Rheum 32, 15-21.

Neumann L, Buskila D (2003): Epidemiology of fibromyalgia. Curr Pain Headache Rep 7, 362-368.

Niemistö L, Rissanen P, Sarna S, Lahtinen-Suopanki T, Lindgren KA, Hurri H (2005): Cost-effectiveness of combined manipulation, stabilizing exercises, and physician consultation compared to physician consultation alone for chronic low back pain: a prospective randomized trial with 2-year follow-up. Spine (Phila Pa 1976) 30, 1109-1115

Nimnuan C, Rabe-Hesketh S, Wessely S, Hotopf M (2001): How many functional somatic syndromes? J Psychosom Res 51, 549-557

Nogid A, Pham DQ (2006): Role of abatacept in the management of rheumatoid arthritis. Clin Ther 28, 1764-1778

O'Brien EM, Staud RM, Hassinger AD, McCulloch RC, Craggs JG, Atchison JW, Price DD, Robinson ME (2010): Patient-centered perspective on treatment outcomes in chronic pain. Pain Med 11, 6-15

O'Connell NE, Wand BM, Marston L, Spencer S, Desouza LH (2011): Non-invasive brain stimulation techniques for chronic pain. A report of a Cochrane systematic review and meta-analysis. Eur J Phys Rehabil Med 47, 309-326

Offenbaecher M, Waltz M, Schoeps P (2000): Validation of a German version of the Fibromyalgia Impact Questionnaire (FIQ-G). J Rheumatol 27, 1984-1988

Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ (2008): Scales to assess the quality of randomized controlled trials: a systematic review. Phys Ther 88, 156-175

Olsen J (1998): Epidemiology deserves better questionnaires. IEA European Questionnaire Group. International Epidemiological Association. Int J Epidemiol 27, 935

Osiri M, Shea B, Robinson V, Suarez-Almazor M, Strand V, Tugwell P, Wells G (2003): Leflunomide for the treatment of rheumatoid arthritis. Cochrane Database Syst Rev 1, CD002047

Osterhaus JT, Purcaru O, Richard L (2009): Discriminant validity, responsiveness and reliability of the rheumatoid arthritis-specific Work Productivity Survey (WPS-RA). Arthritis Res Ther 11, R73

Pach D, Brinkhaus B, Roll S, Wegscheider K, Icke K, Willich SN, Witt CM (2011): Efficacy of injections with Disci/Rhus toxicodendron compositum for chronic low back pain--a randomized placebo-controlled trial. PLoS One 6, e26166

Pauer L, Danneskiold-Samsoe B, Jespersen A (2008): Pregabalin for the management of Fibromyalgia (FM): A 14-week, randomised, double-blind, placebo controlled, monotherapy trial (Study A0081100). Ann Rheum Dis 67(Suppl),256

Pope JE, Prashker M, Anderson J (2004): The efficacy and cost effectiveness of N of 1 studies with diclofenac compared to standard treatment with nonsteroidal antiinflammatory drugs in osteoarthritis. J Rheumatol 31, 140-149

Price C, Arden N, Coglan L, Rogers P (2005): Cost-effectiveness and safety of epidural steroids in the management of sciatica. Health Technol Assess 9, 33

Pye SR, Reid DM, Smith R, Adams JE, Nelson K, Silman AJ, O'Neill TW (2004): Radiographic features of lumbar disc degeneration and self-reported back pain. J Rheumatol 31, 753-758.

Radbruch L, Loick G, Kiencke P, Lindena G, Sabatowski R, Grond S, Lehmann KA, Cleeland CS (1999): Validation of the German version of the Brief Pain Inventory. J Pain Symptom Manage 18, 180-187

Rodgers M, Sowden A, Petticrew M, Arai L, Roberts H, Britten N, Popay J (2009): Testing Methodological Guidance on the Conduct of Narrative Synthesis in Systematic Reviews. Evaluation 15, 49-73

Rodgers M, Epstein D, Bojke L, Yang H, Craig D, Fonseca T, Myers L, Bruce I, Chalmers R, Bujkiewicz S, Lai M, Cooper N, Abrams K, Spiegelhalter D, Sutton A, Sculpher M, Woolacott N (2011): Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. Health Technol Assess 15, 10

Roelofs PD, Deyo RA, Koes BW, Scholten RJ, van Tulder MW (2008): Non-steroidal anti-inflammatory drugs for low back pain. Cochrane Database Syst Rev 1, CD000396

Rogvi-Hansen B, Ellitsgaard N, Funch M, Dall-Jensen M, Prieske J (1991): Low level laser treatment of chondromalacia patellae. Int Orthop 15, 359-361

Rossini M, Di Munno O, Valentini G, Bianchi G, Biasi G, Cacace E, Malesci D, La Montagna G, Viapiana O, Adami S (2007): Double-blind, multicenter trial comparing acetyl l-carnitine with placebo in the treatment of fibromyalgia patients. Clin Exp Rheumatol 25, 182-188

Rubinstein SM, van Middelkoop M, Assendelft WJ, de Boer MR, van Tulder MW (2011): Spinal manipulative therapy for chronic low-back pain: an update of a Cochrane review. Spine (Phila Pa 1976) 36, E825-E846

Rubinstein SM, Terwee CB, Assendelft WJJ, de Boer MR, van Tulder MW (2012) Spinal manipulative therapy for acute low-back pain. Cochrane Database Syst Rev 9, CD008880

Ruiz Garcia V, Jobanputra P, Burls A, Cabello JB, Gálvez Muñoz JG, Saiz Cuenca ES, Fry-Smith A (2011): Certolizumab pegol (CDP870) for rheumatoid arthritis in adults. Cochrane Database Syst Rev 2, CD007649

Schmitz S, Adams R, Walsh CD, Barry M, FitzGerald O (2012): A mixed treatment comparison of the efficacy of anti-TNF agents in rheumatoid arthritis for methotrexate non-responders demonstrates differences between treatments: a Bayesian approach. Ann Rheum Dis 71, 225-230

Schneider H, Maheu E, Cucherat M (2012): Symptom-modifying effect of chondroitin sulfate in knee osteoarthritis: a meta-analysis of randomized placebo-controlled trials performed with structum(®). Open Rheumatol J 6, 183-189

Scott DL (1999): Leflunomide improves quality of life in rheumatoid arthritis. Scand J Rheumatol Suppl. 112, 23-29

Singh JA, Christensen R, Wells GA, Suarez-Almazor ME, Buchbinder R, Lopez-Olivo MA, Tanjong Ghogomu E, Tugwell P (2009) Biologics for rheumatoid arthritis: an overview of Cochrane reviews. Cochrane Database Syst Rev 4, CD007848

Skljarevski V, Ossanna M, Liu-Seifert H, Zhang Q, Chappell A, Iyengar S, Detke M, Backonja M (2009): A double-blind, randomized trial of duloxetine versus placebo in the management of chronic low back pain. Eur J Neurol 16, 1041-1048

Skljarevski V, Desaiah D, Liu-Seifert H, Zhang Q, Chappell AS, Detke MJ, Iyengar S, Atkinson JH, Backonja M (2010a): Efficacy and safety of duloxetine in patients with chronic low back pain. Spine (Phila Pa 1976) 35, E578-E585

Skljarevski V, Zhang S, Desaiah D, Alaka KJ, Palacios S, Miazgowski T, Patrick K (2010b): Duloxetine versus placebo in patients with chronic low back pain: a 12-week, fixed-dose, randomized, double-blind trial. J Pain 11, 1282-1290

Skljarevski V, Zhang S, Chappell AS, Walker DJ, Murray I, Backonja M (2010c): Maintenance of effect of duloxetine in patients with chronic low back pain: a 41-week uncontrolled, dose-blinded study. Pain Med 11, 648-657

Skljarevski V, Zhang S, Iyengar S, D'Souza D, Alaka K, Chappell A, Wernicke J (2011): Efficacy of Duloxetine in Patients with Chronic Pain Conditions. Curr Drug ther 6, 296-303

Skljarevski V, Liu P, Zhang S, Ahl J, Martinez JM (2012): Efficacy and Safety of Duloxetine in Patients with Chronic Low Back Pain Who Used versus Did Not Use Concomitant Nonsteroidal Anti-Inflammatory Drugs or Acetaminophen: A Post Hoc Pooled Analysis of 2 Randomized, Placebo-Controlled Trials. Pain Res Treat 2012, 296710

Smolen J, Landewe R, Mease P, Brzezicki J, Mason D, Luijtens K, van Vollenhoven R, Kavanaugh A, Schiff M, Burmester G, Strand V, Vencovsky J, van der Heijde D (2009): Efficacy and safety of certolizumab pegol plus methotrexate in active rheumatoid arthritis: the RAPID 2 study. Ann Rheum Dis 68, 797– 804.

Smolen JS, Han C, van der Heijde D, Emery P, Bathon JM, Keystone E, Kalden JR, Schiff M, Bala M, Baker D, Han J, Maini RN, St Clair EW (2006): Infliximab treatment maintains employability in patients with early rheumatoid arthritis. Arthritis Rheum 54, 716-722

Sörensen J, Aaro S, Bengtsson M, Kalman S, Reigo T, Tropp H (1996): Can a pharmacological pain analysis in patients with chronic low back pain predict the outcome of lumbar fusion? Preliminary report. Eur Spine J 5, 326-331

Spitzer WO, Skovron ML, Salmi LR, Cassidy JD, Duranceau J, Suissa S, Zeiss E (1995): Scientific monograph of the Quebec Task Force on Whiplash-Associated Disorders: redefining "whiplash" and its management. Spine (Phila Pa 1976) 20 (8 Suppl), 1S-73S

Stacey BR, Glanzman RL (2003): Use of gabapentin for postherpetic neuralgia: results of two randomized, placebo-controlled studies. Clin Ther 25, 2597-608

Stanton-Hicks M, Jänig W, Hassenbusch S, Haddox JD, Boas R, Wilson P (1995): Reflex sympathetic dystrophy: changing concepts and taxonomy. Pain 63, 127-133

Staud R, Vierck CJ, Cannon RL, Mauderli AP, Price DD (2001): Abnormal sensitization and temporal summation of second pain (wind-up) in patients with fibromyalgia syndrome. Pain 91, 165-175

St. Clair EW, van der Heijde D, Smolen J, Maini R, Bathon J, Emery P, Keystone E, Schiff M, Kalden J, Wang B, DeWoody K, Weiss R, Baker D (2004): Combination of infliximab and methotrexate therapy for early rheumatoid arthritis: a randomized, controlled trial. Arthritis Rheum 50, 3432–3443.

Sterne JA, Egger M (2001): Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. J Clin Epidemiol 54, 1046-1055.

Stone DH (1993): Design a questionnaire. BMJ 307, 1264-1266

Strand V, Tugwell P, Bombardier C, Maetzel A, Crawford B, Dorrier C, Thompson A, Wells G (1999): Function and health-related quality of life: results from a randomized controlled trial of leflunomide versus methotrexate or placebo in patients with active rheumatoid arthritis. Leflunomide Rheumatoid Arthritis Investigators Group. Arthritis Rheum 42, 1870-1878

Straube S, Tramèr MR, Moore RA, Derry S, McQuay HJ (2009): Mortality with upper gastrointestinal bleeding and perforation: effects of time and NSAID use. BMC Gastroenterol 9, 41

Straube S, Derry S, Moore RA, Paine J, McQuay HJ (2010): Pregabalin in fibromyalgia--responder analysis from individual patient data. BMC Musculoskelet Disord 11, 150

Straube S, Moore RA, Paine J, Derry S, Phillips CJ, Hallier E, McQuay HJ (2011): Interference with work in fibromyalgia: effect of treatment with pregabalin and relation to pain response. BMC Musculoskelet Disord 12, 125

Suarez-Almazor ME, Belseck E, Shea B, Tugwell P, Wells GA (1998): Methotrexate for treating rheumatoid arthritis. Cochrane Database Syst Rev 2, CD000957

Symmons D, Turner G, Webb R, Asten P, Barrett E, Lunt M, Scott D, Silman A (2002): The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. Rheumatology 41, 793–800

Szczurko O, Cooley K, Mills EJ, Zhou Q, Perri D, Seely D: Szczurko O1, Cooley K, Mills EJ, Zhou Q, Perri D, Seely D (2009): Naturopathic treatment of rotator cuff tendinitis among Canadian postal workers: a randomized controlled trial. Arthritis Rheum 61, 1037-1045

Thompson MS, Read JL, Hutchings HC, Paterson M, Harris ED Jr (1988): The cost effectiveness of auranofin: results of a randomized clinical trial. J Rheumatol 15, 35-42

Tort S, Urrútia G, Nishishinya MB, Walitt B (2012): Monoamine oxidase inhibitors (MAOIs) for fibromyalgia syndrome. Cochrane Database Syst Rev 4, CD009807

Turner JA, Loeser JD, Bell KG (1995): Spinal cord stimulation for chronic low back pain: a systematic literature synthesis. Neurosurgery 37, 1088-1095

van den Bosch MA, Hollingworth W, Kinmonth AL, Dixon AK (2004): Evidence against the use of lumbar spine radiography for low back pain. Clin Radiol 59, 69-76

van der Heijde D, Dijkmans B, Geusens P, Sieper J, DeWoody K, Williamson P, Braun J, Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy Study Group (2005): Efficacy and safety of infliximab in patients with ankylosing spondylitis: results of a randomized, placebo-controlled trial (ASSERT). Arthritis Rheum 52, 582–591

van der Heijde D, Han C, DeVlam K, Burmester G, van den Bosch F, Williamson P, Bala M, Han J, Braun J (2006): Infliximab improves productivity and reduces workday loss in patients with ankylosing spondylitis: results from a randomized, placebo-controlled trial. Arthritis Rheum 55, 569-574

Wells GA, Haguenauer D, Shea B, Suarez-Almazor ME, Welch V, Tugwell P, Peterson J (1998): Cyclosporine for treating rheumatoid arthritis. Cochrane Database Syst Rev 2, CD001083

Wernicke JF, Wang F, Pritchett YL, Smith TR, Raskin J, D'Souza DN, Iyengar S, Chappell AS (2007): An open-label 52-week clinical extension comparing duloxetine with routine care in patients with diabetic peripheral neuropathic pain. Pain Med 8, 503-513

Winemiller MH, Billow RG, Laskowski ER, Harmsen WS (2005): Effect of magnetic vs sham-magnetic insoles on nonspecific foot pain in the workplace: a randomized, double-blind, placebo-controlled trial. Mayo Clin Proc 80, 1138-1145

Wolf I, Friede T, Hallier E, Straube S: Beschreibung arbeitsbezogener Endpunkte in Studien zur Behandlung chronischer Schmerzerkrankungen – systematische Literaturübersicht in: 52.Wissenschaftliche Jahrestagung der Deutschen Gesellschaft für Arbeitsmedizin und Umweltmedizin 14. - 17. März 2012, Göttingen. Dokumentation, 460-463

Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, Tugwell P, Campbell SM, Abeles M, Clark P et al. (1990): The American College of Rheumatology 1990 Criteria for the Classification of Fibromyalgia. Report of the Multicenter Criteria Committee. Arthritis Rheum 33, 160-172

Wood AM, White IR, Thompson SG (2004): Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. Clin Trials 1, 368-376

Woolacott N, Bravo Vergel Y, Hawkins N, Kainth A, Khadjesari Z, Misso K, Light K, Asseburg C, Palmer S, Claxton K, Bruce I, Sculpher M, Riemsma R (2006): Etanercept and infliximab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. Health Technol Assess 10(31), 1-239

Yang H, Craig D, Epstein D, Bojke L, Light K, Bruce IN, Sculpher M, Woolacott N (2012): Golimumab for the treatment of psoriatic arthritis: a NICE single technology appraisal. Pharmacoeconomics 30, 257-270

Zanette Sde A, Born IG, Brenol JC, Xavier RM (2008): A pilot study of acupuncture as adjunctive treatment of rheumatoid arthritis. Clin Rheumatol 27, 627-635

Zhang W, Bansback N, Boonen A, Young A, Singh A, Anis AH (2010): Validity of the work productivity and activity impairment questionnaire--general health version in patients with rheumatoid arthritis. Arthritis Res Ther 12, R177

Zhang Z, Xu X, Ni H (2013): Small studies may overestimate the effect sizes in critical care meta-analyses: a meta-epidemiological study. Crit Care 17, R2

Anhang

Anhang 1

Beschreibung verwendeter Studien

Autor & Jahr	Albert et al. 2013
Erkrankung	Chronische Rückenschmerzen (Osteochondrose)
Probanden	162 randomisiert Teilnahmebedingungen: Alter 18-65 Jahre; mit MRT bestätigter Bandscheibenvorfall bei L3/4, L4/5 oder L5/S1 und korrespondierender Modic 1 Befund ; chronische Rückenschmerzen in der Lumbalregion seit mehr als 6 Wochen Ausschlusskriterien: Allergien gegen Antibiotika; Schwangerschaft/Stillzeit; Nierenerkrankung; bestehende Rechtsstreitigkeiten Probandencharakteristika (randomisiert): Durchschnittsalter in Jahren: 45,1 Anteil weiblicher Probanden: 58,2 % chronische Schmerzen seit mehr als 2 Jahren: 52,4 %
Behandlungen (Interventionen)	Die Patienten erhielten 100 Tage lang entweder 3-mal täglich Amoxicillin-Clavulansäure (500mg/125mg) (n=45), 3-mal täglich 1000mg/259mg (n=45) oder Placebo (36 nahmen 3-mal täglich eine Tablette und weitere 36 3-mal täglich 2 Placebotabletten). Die Patienten durften während der Studie ihre bisherige Behandlung mit Antiphlogistika und Analgetika weiter verfolgen. Die Patienten wurden bei Randomisierung, nach Ablauf der Behandlung (100 Tage) und nach einem Jahr untersucht.
Qualität (OQS)	Punkte: 5/5 Randomisierung: Die Probanden wurden mit einer computergenerierten Zufallssequenz ihren Behandlungen zugeordnet. Punkte: 2/2 Verblindung: Die Studienmedikamente waren optisch identisch und wurden der Behandlung nicht zuordenbar in nummerierten Packungen ausgegeben. Punkte: 2/2 Umgang mit Studienabbrechern: 15 Patienten beendeten die Studie vor Ablauf der Behandlung. 3 weitere Patienten waren bei der Befragung 1 Jahr später nicht mehr verfügbar. Der statistische Umgang mit den fehlenden Daten wurde nicht genauer berichtet. Punkte: 1/1
Studienendpunkte	Primärere Endpunkte: - Roland-Morris Disability Questionnaire-24 (RMDQ-24) - Low Back Pain Rating Scale Sekundäre Endpunkte: - Global Perceived Effect (GPE) - Stunden mit Rückenschmerzen während der letzten 4 Wochen - EuroQoL: 5 Dimensions Questionnaire (EQ-5D) - verlorene Arbeitstage - „Bothersomeness“ nach Dunn - MRT Modic Grading - Serumanalyse

Arbeitsbezogene Endpunkte	Arbeitsfehltage im vergangenen Jahr; erhoben bei Studienbeginn, sowie 1 Jahr später. Statistische Methoden sind nicht zu zuordnen.

Autor & Jahr	Arnold et al. 2005
Erkrankung	Fibromyalgie
Probanden	354 randomisiert Teilnahmebedingungen: Alter ≥ 18 Jahre; weibliches Geschlecht; bestätigte Diagnose auf Fibromyalgie nach den ACR Kriterien; Durchschnittsschmerz ≥ 4/10 (BPI) Ausschlusskriterien: Schmerzen von Verletzungen oder entzündlichen rheumatischen Erkrankungen; schwere Allgemeinerkrankung; Vorliegen psychischer Erkrankungen, ausgenommen schwere depressive Episode oder Angstzustände; bekannter Drogenmissbrauch; suizidgefährdeter Patient; Schwangerschaft/Stillzeit; Probanden, die nach Meinung des Untersuchers behandlungsresistent sind; vorhergehende Teilnahme an einer Duloxetin-Studie; Beantragung von die Krankheit betreffenden sozialen Aufwendungen; allergische Reaktionen gegen viele verschiedene Medikamente Probandencharakteristika: Durchschnittsalter in Jahren: 49,6 Anteil weiblicher Probanden: 100% durchschnittliche Jahre seit Diagnose: nicht berichtet
Behandlungen (Interventionen)	Die Patienten erhielten 12 Wochen lang entweder 60 mg/Tag Duloxetin (n=118), 120 mg/Tag Duloxetin (n=116) oder Placebo (n=120). Bis auf Paracetamol (max. 2 g/Tag) zu analgetischen Zwecken und ASS (max. 325 mg/Tag) zur Antikoagulation waren andere Therapien nicht zugelassen.
Qualität (OQS)	Punkte: 5/5 Randomisierung: Das Verfahren der Randomisierung wurde nicht berichtet. Stratifiziert wurde nach Vorliegen einer schweren depressiven Episode. Punkte: 2/2 Verblindung: Es wurde nicht näher beschrieben wie Doppel-Verblindung gewahrt wurde. Punkte: 1/2 Umgang mit Studienabbrechern: 139 Probanden beendeten die Behandlung vorzeitig. Diese wurden in die Analyse einbezogen, wenn mindestens ein Datensatz nach Baseline erhoben wurde. Dies lässt auf eine LOCF-Herangehensweise schließen. Sekundär wurde ebenfalls ein MMRM (*mixed model for repeated measurements*) zur Analyse der longitudinalen Entwicklung während der Studie angefertigt. Punkte: 1/1
Studienendpunkte	Primärer Endpunkte: - Brief Pain Inventory (BPI) 24h-Durchschnittsschmerz Sekundäre Endpunkte: - BPI Schmerzen (Durchschnitt, schlimmster, momentan) - BPI Interferenz - Fibromyalgia Impact Questionnaire (FIQ) - Clinical Global Impressions of Severity (CGI-S) - Patient's Global Impression of Improvement (PGI-I) - Hamilton Depression Rating Scale (HAMD)
Arbeitsbezogene Endpunkte	BPI Unterpunkt: Interferenz mit normaler Arbeit. Eine Kovarianzanalyse wurde durchgeführt. Kovarianten: Behandlung, Untersucher und Baseline-Werte

Autor & Jahr	Arnold et al. 2009 Bradley et al. 2010
Erkrankung	Fibromyalgie
Probanden	1332 analysiert Teilnahmebedingungen: Geschlecht (in 3 Studien männlich und weiblich, 1 Studie untersuchte nur weibliche

	Patienten); Alter ≥ 18 Jahre; bestätigte Diagnose nach ACR Ausschlusskriterien: Schwere physische Erkrankung; eine andere psychische Erkrankung als schwere depressive Episode; andere schmerzhafte Erkrankungen Probandencharakteristika der analysierten Patienten: Durchschnittsalter in Jahren: 50,2 Anteil weiblicher Probanden: 94,7 % durchschnittliche Jahre seit Diagnose: nicht berichtet
Behandlungen (Interventionen)	In den untersuchten Studien erhielten die Patienten der aktiven Gruppe 20, 60 oder 120 mg/Tag Duloxetin. Zwei der Studien wurden über 12 Wochen, die beiden anderen über 28 Wochen hinweg durchgeführt..
Qualität (OQS)	Meta-Analyse
Studienendpunkte	- Fibromyalgia Impact Questionnaire (FIQ) - Short Form 36 Health Survey Questionnaire (SF-36) - Brief Pain Inventory (BPI) - Patient's Global Impression of Improvement (PGI-I) - Sheehan Disability Scale (SDS)
Arbeitsbezogene Endpunkte	- FIQ: Interferenz mit normaler Arbeit - SDS Beeinträchtigung bei der Arbeit oder in der Schule Fehlende Daten wurden mit der LOCF-Methodik verarbeitet.

Autor & Jahr	Barkham et al. 2010
Erkrankung	Spondylitis ankylosans
Probanden	40 randomisiert Teilnahmebedingungen: bestehende Spondylitis ankylosans nach den Definitionen der modifizierten New York Kriterien; mindestens einen Score von 40 Punkten des „Bath Ankylosing Spondylitis Disease Activity Index" (BASDAI); VAS-pain ≥ 40/100; Morgensteifigkeit welche mindestens 45 min anhält; Probanden mussten ein Beschäftigungsverhältnis vorweisen, aber auch auf der Work Instability Scale einen Wert größer 10 erreichen (definiert als instabile Prognose der Arbeitsfähigkeit) Ausschlusskriterien: durchgemachte oder aktuelle Tuberkuloseinfektion; kongestive Herzinsuffizienz; Behandlung mit Glucocorticoiden im vergangenen Monat Probandencharakteristika: Durchschnittsalter in Jahren: 40,1 Anteil weiblicher Probanden: 20 % durchschnittliche Jahre seit Diagnose: 15,5
Behandlungen (Interventionen)	Die Patienten erhielten 12 Wochen lang 2-mal wöchentlich 25mg Entanercept (subkutan) oder Placebo. Erhielten die Patienten bereits vor der Studie regelmäßig Medikamente, so durften diese unter Beibehaltung der Dosis weiter eingenommen werden. Erlaubt waren Sulfasalazin oder Methotrexat in Kombination mit NSARs.
Qualität (OQS)	Punkte: 3/5 Randomisierung: nicht beschrieben Punkte: 1/2 Verblindung: Es wurde berichtet, dass Untersuchungspersonal, Forscher-Team und Patienten bis zum Ende der Studie über die Behandlung bei einzelnen Individuen im Unklaren waren. Eine Genauere Beschreibung der Methodik fehlt. Punkte: 1/2 Umgang mit Studienabbrechern: Keiner der untersuchten Patienten beendeten die Studie vorzeitig. Punkte: 1/1
Studienendpunkte	Primäre Endpunkte: - WIS Sekundäre Endpunkte: - BASDAI - Ankylosing Spondylitis Quality of Life (ASQoL)

	- Bath Ankylosing Spondylitis Functional Index (BASFI) - Disability Index of Stanford Health Assessment Questionnaire (HAQ-DI) - verlorene Arbeitsstunden (Arbeitstagebuch)
Arbeitsbezogene Endpunkte	- WIS: die ordinalen Daten wurden mit einem Rash-Modell in eine Intervallskala transformiert und mit einer Kovarianzanalyse verglichen, Kovarianten: Baselinedaten; die kategorische Einteilung der WIS (geringes, mittleres und hohes Risiko des Arbeitsplatzverlustes) wurden mit dem Mann-Whitney U Test verglichen. - Verlorene Arbeitstage während der Studie: Es konnten nur Arbeitstagebücher von 25 Patienten ausgewertet werden.

Autor & Jahr	Baron et al. 2010
Erkrankung	Chronische Rückenschmerzen aufgrund von spinaler Stenose oder Bandscheibenprolaps
Probanden	217 randomisiert Teilnahmebedingungen: Alter ≥ 18 Jahre; Schmerzen aufgrund von spinaler Stenose oder Bandscheibenprolaps für mindestens 3 Monate aber maximal für 4 Jahre; wöchentlicher Durchschnittsschmerz mindestens 4/10 Punkte nach der Vorlaufphase Ausschlusskriterien: lumbosakrale Operation während des letzten halben Jahres; mehr als ein lumbosakraler Eingriff; epidurale Injektionen während der letzten 6 Wochen Probandencharakteristika der randomisierten Patienten: Durchschnittsalter in Jahren: 52,5 Anteil weiblicher Probanden: 52,1 % durchschnittliche Jahre seit Diagnose: nicht berichtet
Behandlungen (Interventionen)	Nach einer Vorlaufphase, in der eine Woche lang alle Patienten Placebos erhielten, und einer vierwöchigen Einfachblindphase, in der alle Patienten eine Dosis von 150-600 mg/Tag Pregabalin erhielten, wurden 159 Patienten von der Studie ausgeschlossen, da diese Patienten positiv auf Placebo reagierten (über 50% Schmerzreduktion), nicht auf Pregabalin reagierten(weniger als 30% Schmerzreduktion), oder sie wurden wegen Nebenwirkungen oder anderen Gründen aus der Studie entfernt. 217 Patienten wurden anschließend über 5 Wochen hinweg entweder mit Placebo (n = 107) oder mit Pregabalin (n = 110) behandelt. Die individuellen Dosen wurden in der einfachverblindeten Phase entschieden. Es durften keine Antikonvulsiva, hochpotente Opioide oder Opioid-Kombinationsmedikamente genommen werden. Erhielten die Patienten bereits vor der Studie regelmäßig Medikamente, so durften diese unter Beibehaltung der Dosis weiter eingenommen werden. Während der doppelblinden Phase durfte keine Notfallmedikation eingenommen werden.
Qualität (OQS)	Punkte: 4/5 Randomisierung: Es wurde eine Zufallssequenz verwendet, die an ein Internet/Telefon-Interface gekoppelt war. Punkte: 2/2 Verblindung: Es wurde nichts über die Beschaffenheit der Medikamente berichtet. Punkte: 1/2 Umgang mit Studienabbrechern: Nach dem Beginn der doppelblinden Phase brachen 30 Patienten die Studie vorzeitig ab. Über den statistischen Umgang mit fehlenden Werten wurde nicht berichtet. Punkte: 1/1
Studienendpunkte	Primärer Endpunkte: - Daily Pain Rating Scale scores (DPRS) - Verlust des Ansprechens auf die Behandlung (Anstieg von ≥ 1 Punkt (DPRS) verglichen zum Ausgangswert, Verwendung von Rettungsmedikamenten) oder Studienabbruch Sekundäre Endpunkte: - Schlafstörungen - Hospital Anxiety and Depression Scale (HADS) - Generalized Anxiety Visual Analogue Scale (GA-VAS) - Patient Global Impression of Change (PGIC) - Medical Outcomes Study (MOS) - Pain Treatment Satisfaction Scale - Roland-Morris Disability Scale (RDQ)

	- Work Productivity and Activity Impairment Questionnaire (WPAI) - EuroQoL: 5 Dimensions Questionnaire (EQ-5D) - Lumbosacral Radiculopathy Pain Management Questionnaire (LRPMQ)
Arbeitsbezogene Endpunkte	Work Productivity and Activity Impairment Questionnaire (WPAI) Es wurde eine Varianzanalyse durchgeführt. Da keine signifikanten Unterschiede zwischen der aktiven Gruppe und der Kontrollgruppe festgestellt werden konnte, wurde auf eine genauere Beschreibung verzichtet.

Autor & Jahr	Bejarano et al. 2008
Erkrankung	Rheumatoide Arthritis
Probanden	148 randomisiert Teilnahmebedingungen: Alter ≥ 18 Jahre; bestätigte Diagnose nach ACR; weniger als 2 Jahre Symptomdauer; Vorliegen einer bezahlten Beschäftigung; krankheitsbedingte Einschränkungen während der Arbeit; keine vorhergehende Behandlung mit Methotrexat Ausschlusskriterien: Infektionserkrankungen; durchgemachte bösartige Tumorerkrankung; schwere Allgemeinerkrankung; Schwangerschaft/Stillzeit; dauerhafte Krankschreibung; Demyelinisierung; andere entzündliche rheumatische Erkrankungen Probandencharakteristika: Durchschnittsalter in Jahren: 47 Anteil weiblicher Probanden: 56 % durchschnittliche Monate seit Diagnose: 8,7
Behandlungen (Interventionen)	Alle 2 Wochen erhielten die Patienten subkutan 56 Wochen lang entweder Adalimumab (n=75) oder Placebo (n=73). Die Dosis von 7,5 mg/Woche wurde auf 25 mg erhöht, falls in Woche 12 weiterhin Synovitis festgestellt werden konnte. Zusätzlich wurde Folsäure (5mg 6- mal die Woche) gegeben. Geschwollene Gelenke durften mittels intraartikulärer Injektion von Methyl-Prednisolon behandelt werden, aber nur 80 mg insgesamt während der gesamten 56 Wochen. Stabile Dosen von Prednisolon (max. 10 mg/Tag) oder anderen Analgetika durften weiter eingenommen werden. Die Patienten schieden aus der Studie aus, wenn sie entweder ihre Anstellung verloren oder Verlust des Arbeitsplatzes drohte, welches als Nichterreichen des ACR20 und Verschlechterung oder Beibehaltung eines höheren Wertes als 17 Punkte auf der WIS-Skala in oder nach Woche 16 definiert war.
Qualität (OQS)	Punkte: 5/5 Randomisierung: Es wurde eine zentrale computergenerierte Randomisierung durchgeführt. Punkte: 2/2 Verblindung: Im Randomisierungszentrum wurden die nummerierten Pakete mit den Lösungen vorbereitet und an die Behandlungsorte geschickt. Die Adalimumab- und Placebolösungen waren visuell identisch. Punkte: 2/2 Umgang mit Studienabbrechern: 62 Patienten beendeten die Studie vorzeitig. Für den primären Endpunkt (Arbeitsplatzverlust/drohender Arbeitsplatzverlust nach Woche 16) wurden alle Studienabbrecher die wegen anderer Gründe als Arbeitsplatzverlust bzw. drohende Arbeitslosigkeit ausschieden als noch in der Studie anwesend und arbeitend behandelt. Verglichen wurden diese Ergebnisse mit Untersuchungen in denen diese Abbrecher als wegen Arbeitslosigkeit ausgeschiedene behandelt wurden. Punkte: 1/1
Studienendpunkte	Primärer Endpunkte: - Arbeitsplatzverlust/drohender Arbeitsplatzverlust nach Woche 16 sekundäre Endpunkte: - Arbeitsplatzverlust/drohender Arbeitsplatzverlust nach Woche 56 - Disease Activity Score in 28 joints (DAS28) - American College of Rheumatology 20% response (ACR20) - Work Instability Scale (WIS) - Health Assessment Questionnaire (HAQ) - Rheumatoid Arthritis Quality of Life questionnaire (RAQoL) - verlorene Arbeitszeit in Prozent

Arbeitsbezogene Endpunkte	- Arbeitsplatzverlust/drohender Arbeitsplatzverlust nach 16 bzw. 56 Wochen, dargestellt mit einer Chi-Quadrat-Statistik sowie einer Kaplan-Meier-Kurve. Für Studienabbrecher wurden verschiedene Annahmen ausgewertet: > Abbrecher behandelt als im Beruf verbleibend > Abbrecher behandelt als arbeitslos > LOCF > Patienten welche aufgrund hoher Krankheitsaktivität die Studie verließen wurden als arbeitslos gewertet - Zusätzliche Untergruppenanalysen für auf nur Rheumatoide Arthritis zurückzuführenden Arbeitsplatzverlust sowie für nur drohenden Arbeitsplatzverlust wurden durchgeführt. - Verlorene Arbeitszeit während der Studie in Prozent - Work Instability Scale: Es wurde eine Varianzanalyse durchgeführt. Umgang mit fehlenden Daten wurde nicht beschrieben.

Autor & Jahr	Bennett et al. 2003 & 2005
Erkrankung	Fibromyalgie
Probanden	315 randomisiert Teilnahmebedingungen: Alter 18-75 Jahre; Schmerzempfinden berichtet als mindestens 40/100 mm (VAS); bestätigte Diagnose auf Fibromyalgie nach den Kriterien des ACR Ausschlusskriterien: schwere Allgemeinerkrankung oder andere relevante Schmerzerkrankung; keine Kontrazeption; vorherige erfolglose Behandlung mit Tramadol; Behandlung mit Tramadol während des letzten Monats; Behandlung mit Antikonvulsiva, Akupunktur, Cyclobenzaprin, Antidepressiva, TENS während der letzten 3 Wochen; vorherige Behandlung mit Sedativa, Analgetika, Anästhetika, Muskelrelaxantien (oral/topisch) innerhalb ihrer 5fachen Halbwertszeit; Trigger-Punkt-Injektionen während der letzten 2 Monate; systemische Steroid-Gabe während der letzten 3 Monate; Einnahme eines anderen Studienmedikamentes während des vorhergehenden Monats Probandencharakteristika der analysierten Patienten: Durchschnittsalter in Jahren: 50,0 Anteil weiblicher Probanden: 93,9 % durchschnittliche Jahre seit Diagnose: nicht berichtet
Behandlungen (Interventionen)	Die Probanden erhielten 2 Monate lang entweder Paracetamol-Tramadol (37,5 mg/325 mg) (n=158) oder Placebo-Tabletten (n=157). Pro Tag wurden 4-8 Tabletten eingenommen. Nahmen die Patienten bereits einen Monat vor Studienbeginn gleichförmige Dosen von Serotonin-Wiederaufnahmehemmern oder Johanniskraut-Medikamente, so durften diese weiter eingenommen werden. Ebenfalls durften seit einem Monat etablierte Schlafmittel (Zolpidem, Flurazepam) weiter verwendet werden.
Qualität (OQS)	Punkte: 3/5 Randomisierung: Die Patienten wurden mit einem randomisierten Behandlungszeitplan ihren Medikamenten zugeordnet. Die Art der verwendeten Randomisierung wurde nicht näher beschrieben. Punkte: 1/2 Verblindung: Die Behälter der Tabletten waren mit einem Schildchen überklebt, welches im Notfall entfernt werden konnte, um die Verblindung zu beenden. Über die Beschaffenheit der verwendeten Medikamente wurde nicht berichtet. Punkte: 1/2 Umgang mit Studienabbrechern: 2 Patienten wurden wegen Nichterscheinens bei wenigstens einer Post-Basline-Untersuchung aus der Analyse entfernt. 175 Patienten beendeten die Studie vorzeitig. Der statistische Umgang mit fehlenden Daten wurden nicht beschrieben. Punkte: 1/1
Studienendpunkte	- Vergleich der Baseline-Werte mit genereller US-Population unter Verwendung von > Health-related Quality of Life (HRQOL) > Medical Outcomes Study (MOS) Unterpunkt: Schlaf - Einfluss der Behandlung unter Verwendung von: > Health-related Quality of Life (HRQOL) > Fibromyalgia Impact Questionnaire (FIQ) > Short Form 36 Health Survey Questionnaire (SF-36) > Arbeitszeitverluste der letzten 7 Tage (FIQ)
Arbeitsbezogene	- FIQ Unterpunkt: Beeinträchtigungen bei der Arbeit und Arbeitszeitverluste; Kovarianzanalyse wurde

Endpunkte	durchgeführt - SF-36 Unterpunkt: Interferenz mit Arbeitsablauf wurde nur für die aktive Gruppe berichtet. Es wurde kein Vergleich mit der Placebogruppe erarbeitet.

Autor & Jahr	Carlsson und Sjölund 2001
Erkrankung	Chronische Rückenschmerzen
Probanden	51 randomisiert (davon 50 analysiert) Teilnahmebedingungen: Schmerzen des unteren Rückens in lumbal oder lumbosakral Region während der letzten 6 Monate oder länger; Ausstrahlung der Schmerzen nur bis in das proximale Glied der unteren Extremitäten; normale neurologische Befunde der unteren spinalen Nerven (afferent und efferent) Ausschlusskriterien: schwere Allgemeinerkrankung oder Trauma; Schwangerschaft; vorhergehende Erfahrung mit Akupunktur Probandencharakteristika: Durchschnittsalter in Jahren: 49,8 Anteil weiblicher Probanden: 66 % durchschnittliche Jahre seit Diagnose: 9,5
Behandlungen (Interventionen)	Die Patienten wurden entweder mit manueller Akupunktur (n=18), Elektroakupunktur (n=16) oder einer Placebo-TENS (n=16) behandelt. In den ersten 2 Monaten wurde jede Woche eine Behandlung durchgeführt. 2 nachfolgende Behandlungen wurden im 4. und 6. Monat durchgeführt. Die Patienten wurden bis in den 8. Monat untersucht. Patienten die im 3. Monat keine Verbesserung zeigten wurden aus der Studie entfernt. Die Patienten durften in Selbstmedikation NSARs zu sich nehmen.
Qualität (OQS)	Punkte: 5/5 Randomisierung: Wurde mit einer computergenerierten Zufallssequenz durchgeführt. Punkte: 2/2 Verblindung: Die Zufallssequenz wurde von unabhängigem Klinikpersonal aufbewahrt. Das Placebogerät war so beschaffen, dass es bei Aktivierung blinkte und so eine Intervention vortäuschte. Patientenbefragungen und die Auswertungen der Patiententagebücher wurden von unabhängigem Personal durchgeführt, welches keine Kenntnis von der durchgeführten Therapie hatte. Punkte: 2/2 Umgang mit Studienabbrechern: Von den 51 behandelten Patienten erschien einer nicht beim Wiedervorstellungstermin und wurde von der Analyse ausgeschlossen. Alle anderen fehlenden Daten wurden mit der LOCF-Methodik behandelt. Punkte: 1/1
Studienendpunkte	- Generelle Veränderung (basierend auf Patienten-Interviews): schlechter, unverändert, verbessert - Visual Analogue Scale bezüglich des Schmerzes morgens und abends - Beschäftigungsstatus (Vollzeit, Halbtags, krankgeschrieben) - Schlafstörungen - Analgetikakonsum
Arbeitsbezogene Endpunkte	Der Beschäftigungsstatus wurde von 32 Patienten untersucht, die restlichen Probanden waren bereits verrentet. Die Veränderungen innerhalb der Gruppe gegenüber Baseline wurden mit dem Wilcoxon-Vorzeichen-Rang-Test berechnet.

Autor & Jahr	Chappell et al. 2008
Erkrankung	Fibromyalgie
Probanden	330 randomisiert Teilnahmebedingungen: Alter ≥ 18 Jahre; Probanden erfüllten die Kriterien des ACR für Fibromyalgie mit oder ohne Vorliegen einer schweren depressiven Episode Ausschlusskriterien: vorhergehende Therapie mit Duloxetin; andere psychische Erkrankung als Depression; verletzungsbedingte Schmerzerkrankungen; entzündliche oder gewebsspezifische lokale rheumatische

	Erkrankungen; lokal begrenzte Schmerzsyndrome; schwere anderweitige Erkrankung Probandencharakteristika: Durchschnittsalter in Jahren: 50,5 Anteil weiblicher Probanden: 93,3 % durchschnittliche Jahre seit Diagnose: nicht berichtet
Behandlungen (Interventionen)	Die Probanden erhielten 27 Wochen lang entweder Placebo (n=168) oder 60 mg/Tag Duloxetin (n=162). Konnte der BPI Durchschnittsschmerz in Woche 13 nicht um mindestens 50% verbessert werden, so wurde die Dosis auf 120 mg/Tag erhöht. Wurde die Dosiserhöhung nicht vertragen, so konnte sie auf 60 mg zurück gefahren werden.
Qualität (OQS)	Punkte: 3/5 Randomisierung: Das Randomisierungsverfahren wurde nicht näher beschrieben. Punkte: 1/2 Verblindung: Die Methode mit der Doppelblindheit erreicht wurde, wurde nicht beschrieben. Es wurde jedoch berichtet, dass die Dosiserhöhung auch mit Bedacht auf die Wahrung der Doppelblindheit organisiert wurde. Punkte: 1/2 Umgang mit Studienabbrechern: 126 Patienten beendeten die Studie vorzeitig. Signifikant mehr Patienten verließen die Placebo-Gruppe wegen mangelnder Wirkung als in der aktiven Gruppe. Für die Kovarianzanalysen wurden die letzten verfügbaren Daten verwendet. Punkte: 1/1
Studienendpunkte	Primäre Endpunkte: - BPI Durchschnittsschmerz (Baseline-Endpunkt) - Patient's Global Impressions of Improvement questionnaire (PGI-I) Sekundäre Endpunkte: - Fibromyalgia Impact Questionnaire (FIQ) - Clinical Global Impressions of Severity (CGI-Severity) - Schmerzschwelle der 18 Trigger-Punkte - BPI: verschiedene Unterpunkte - Multidimensional Fatigue Inventory (MFI) - Hamilton Depression Rating Scale (HAMD) - Beck Depression Inventory-II (BDI-II) - Sheehan Disability Scale (SDS) - Short Form 36 Health Survey Questionnaire (SF-36) - EuroQoL: 5 Dimensions Questionnaire (EQ-5D)
Arbeitsbezogene Endpunkte	SDS: Beeinträchtigung bei der Arbeit oder in der Schule Eine Kovarianzanalyse wurde durchgeführt mit LOCF-Daten und mit Baselinedaten als Kovarianten.

Autor & Jahr	Chappell et al. 2009
Erkrankung	Osteoarthritis des Knies
Probanden	231 randomisiert Teilnahmebedingungen: Alter ≥ 40 Jahre; Probanden erfüllten die Kriterien des ACR für Osteoarthritis im Knie, hatten an mehr als der Hälfte der Tage der letzten 3 Monate Schmerzen; der tägliche Durchschnittsschmerz der letzten Woche vor Randomisierung musste ≥ 4/10 betragen Ausschlusskriterien: BMI > 40 kg/m$^{2;}$ Diagnose einer entzündlichen Arthritis oder einer Autoimmunerkrankung; Operationen in den letzten 3 Monaten oder Gelenkprothesen am untersuchten Knie; Vorliegen einer anderen Schmerz verursachenden Erkrankung; Kniearthroskopie während des letzten Monats; Knieoperationen während der letzten 3 Monate; künstliches Kniegelenk; Benutzung einer Gehhilfe; schwere psychische oder allgemeine Erkrankungen; Schwangerschaft/Stillzeit; aktueller oder zurückliegender Drogenmissbrauch; vorhergehende Behandlung mit Duloxetin; Anfallsleiden; unbehandeltes Engwinkelglaukom; akute Leberschäden/Leberzirrhose; schwerwiegende Allergien gegen zahlreiche Medikamente; sowie gegen Duloxetin Probandencharakteristika: Durchschnittsalter in Jahren: 62,3 Anteil weiblicher Probanden: 65,4 % durchschnittliche Jahre seit Diagnose: 7,0

Behandlungen (Interventionen)	Die Patienten erhielten 13 Wochen lang entweder Duloxetin (n=111) oder Placebo (n=120). Die Duloxetindosis betrug in der ersten Woche 30 mg/Tag und wurde dann auf 60 mg/Tag angehoben. In Woche 7 wurde die aktive Gruppe noch einmal randomisiert. So erhielten 43 Patienten nun eine Dosis von 120 mg/Tag, während die restlichen 46 weiterhin 30 mg/Tag bekamen. Patienten durften kurz wirkende Analgetika bei akuten Erkrankungen oder als Rettungsmedikation höchstens an 3 aufeinander folgenden Tagen nehmen, aber insgesamt maximal an 20 Tagen. Patienten, die bereits ein NSAR oder Paracetamol nahmen, durften diese Medikation fortsetzen. Die Dosierung durfte verringert, aber nicht erhöht werden.
Qualität (OQS)	Punkte: 5/5 Randomisierung: Zur Randomisierung wurde eine computergenerierte Zufallssequenz verwendet (stratifiziert nach NSAR/Paracetamol Einnahme). Punkte: 2/2 Verblindung: Zur Wahrung der Doppel-Verblindung wurde die Zufallssequenz mit einem Sprachdialogsystem ausgegeben. Die Placebos sowie die verschiedenen Dosierungen der aktiven Gruppen waren nicht zu unterscheiden. Sie glichen sich in Aussehen, Geruch und Geschmack. Alle Probanden erhielten die gleiche Anzahl an Pillen. Punkte: 2/2 Umgang mit Studienabbrechern: 58 Probanden beendeten die Studie vorzeitig. Zwischen den Gruppen gab es keinen signifikanten Unterschied bezüglich des Abbruchgrundes. Der Primäre Endpunkte wurde mit einer MMRM-Analyse untersucht. Die Kovarianzanalyse der sekundären Endpunkte wurden mit den letzten erfassten Messdaten durchgeführt. Punkte: 1/1
Studienendpunkte	Primäre Endpunkte: - 24h-Durchschnittsschmerz auf einer Linkert Skala (0 für keinen Schmerz, 10 für schlimmster, vorstellbarer Schmerz) Sekundäre Endpunkte: - Patient's Global Impression of Improvement (PGI-I) - Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) - BPI-S (Severity), BPI-I (Interference) - Clinical Global Impressions of Severity (CGI-S) - EuroQoL: 5 Dimensions Questionnaire (EQ-5D) Version des „European Quality of Life instrument" - Beck Depression Inventory-II (BDI-II) - Hospital Anxiety and Depression Scale anxiety subscale (HADS-A)
Arbeitsbezogene Endpunkte	BPI-I: Unterpunkt Interferenz mit normaler Arbeit Eine Kovarianzanalyse wurde durchgeführt mit LOCF Endpunkten; Kovarinaten: Baselinedaten; stratifiziert nach NSAR/Paracetamol Einnahme

Autor & Jahr	Chappell et al. 2011
Erkrankung	Osteoarthritis des Knies
Probanden	256 randomisiert Teilnahmebedingungen: Alter ≥ 40 Jahre; Probanden erfüllten die Kriterien des ACR für Osteoarthritis im Knie und hatten an mehr als der Hälfte der Tage der letzten 3 Monate Schmerzen; der tägliche Durchschnittsschmerz der letzten Woche vor Randomisierung musste ≥ 4/10 betragen Ausschlusskriterien: BMI > 40 kg/m^2; Diagnose einer entzündlichen Arthritis oder einer Autoimmunerkrankung; Operationen in den letzten 3 Monaten oder Gelenkprothesen an untersuchtem Knie; Benutzung einer Gehhilfe; psychische Erkrankungen; Schwangerschaft/Stillzeit; aktueller oder zurückliegender Drogenmissbrauch; reduzierter Allgemeinzustand; Anfallsleiden; unbehandeltes Engwinkelglaukom; akute Leberschäden/Leberzirrhose; schwerwiegende Allergien gegen zahlreiche Medikamente sowie gegen Duloxetin; vorhergehende Behandlung mit Duloxetin Probandencharakteristika: Durchschnittsalter in Jahren: 62,6 Anteil weiblicher Probanden: 76,6%

	durchschnittliche Jahre seit Diagnose: 5,9
Behandlungen (Interventionen)	Die Probanden wurden 13 Wochen lang einmal täglich entweder mit Placebo (n=128) oder mit Duloxetin (n=128) behandelt. In den ersten 6 Wochen erhielten sie 60 mg. War der BPI 24h Durchschnittsschmerz im Vergleich zum Ausgangswert nicht um 30% gesunken, so erhielten sie die kommenden 7 Wochen 120 mg oder abgewandeltes Placebo. Patienten durften kurz wirkende Analgetika bei akuten Erkrankungen oder rheumatischen Rückfällen höchstens an 3 aufeinander folgenden Tagen nehmen, aber insgesamt maximal an 20 Tagen. Patienten die bereits ein NSAR oder Paracetamol nahmen, durften diese Medikation fortsetzen, solange die Dosierung nicht erhöht wurde.
Qualität (OQS)	Punkte: 5/5 Randomisierung: Verwendung eines computergenerierten Zufallsalgorithmus, sowie Stratifizierung unter Einbeziehung von NSAR-Einnahme. Punkte: 2/2 Verblindung: Zur Wahrung der Doppelblindheit wurde ein Sprachdialogsystem gekoppelt an den Zufallsalgorithmus verwendet. Bei der Dosiserhöhung wurden ebenfalls die Placebodosierungen erhöht. Punkte: 2/2 Umgang mit Studienabbrechern: 52 Patienten beendeten die Studie vorzeitig. Die Duloxetin-Gruppe verlor signifikant mehr Probanden wegen Nebenwirkungen als die Placebo-Gruppe. Für die Untersuchung des primären Endpunktes wurden mehrere Herangehensweisen in Bezug auf die Verarbeitung der Studienabbrecher verwendet. Es wurde ein MMRM (*mixed effects model repeated measures*; ein gemischtes Modell sich wiederholender Messungen mit Maximum-Likelihood-Schätzern) berechnet. Eine Kovarianzanalyse mit der BOCF sowie der LOCF Herangehensweisen wurde durchgeführt. Zusätzlich wurde eine modifizierte Version der BOCF Methodik verwendet, bei der nur bei Patienten, die als Studienaustrittsgründe Nebenwirkungen oder fehlende Wirkung nannten, mit Baseline-Daten gerechnet wurde, während alle anderen Probanden mit der LOCF Methode einflossen. Die sekundären Endpunkte wurden ausschließlich mit einem LOCF ANCOVA-Modell berichtet. Punkte: 1/1
Studienendpunkte	Primäre Endpunkte: - Brief Pain Inventory (Unterpunkt 24h-Durchschnittsschmerz) Sekundäre Endpunkte: - Patient's Global Impression of Improvement (PGI-I) - Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) - BPI-S (Severity), BPI-I (Interference) - Short Form 36 Health Survey Questionnaire (SF-36) - Clinical Global Impressions of Severity (CGI-S) - EuroQoL: 5 Dimensions Questionnaire (EQ-5D) Version des „European Quality of Life instrument" - Beck Depression Inventory-II (BDI-II) (nicht berichtet) - Hospital Anxiety and Depression Scale anxiety subscale (HADS-A) (nicht berichtet)
Arbeitsbezogene Endpunkte	BPI-I: Unterpunkt Interferenz mit normalem Arbeitsablauf Eine Kovarianzanalyse wurde durchgeführt mit Baselinewerten und LOCF Endpunkten, stratifiziert nach NSAR Einnahme.

Autor & Jahr	Egsmose et al. 1997
Erkrankung	Reaktive Arthritis
Probanden	83 randomisiert (davon 79 analysiert) Teilnahmebedingungen: Alter > 18 Jahre; bestehende reaktive Arthritis definiert als mindestens ein geschwollenes peripheres Gelenk während NSAR-Behandlung Ausschlusskriterien: Andere rheumatische oder pararheumatische Erkrankung (rheumatoide Arthritis, Spondylitis ankylosans, Psoriasis, chronisch-entzündlichen Darmerkrankungen, akute intermittierende Porphyrie sowie Gelenkserkrankungen aufgrund von Salzausfällungen oder Infektionen); Glucose-6-Phosphat-Dehydrogenase-Mangel; bekannte Allergien gegen Salicylsäure oder Sulfonamide; Neutropenie (< 1500/µl);

	Thrombopenie (< 100000/µl); Leber- oder Niereninsuffizienz; orale oder intraartikuläre Kortikosteroidgabe während der letzten 4 Wochen; Schwangerschaft/Stillzeit; vorhergehende Behandlung mit Sulfasalazin Probandencharakteristika der analysierten Patienten: Durchschnittsalter in Jahren: 34,2 Anteil weiblicher Probanden: 26,6 % durchschnittliche Jahre seit Diagnose: 4,3
Behandlungen (Interventionen)	Die Patienten bekamen, nach dreiwöchiger Eskalation, über 6 Monate hinweg zweimal täglich 1500mg Sulfasalazin (n=37) oder Placebo (n=42).
Qualität (OQS)	Punkte: 3/5 Randomisierung: Die Randomisierung wurde nicht genauer beschrieben. Punkte: 1/2 Verblindung: Über die Erscheinung der Medikamente oder anderweitige Maßnahmen zur Wahrung der Doppelblindheit wurde nicht berichtet. Punkte: 1/2 Umgang mit Studienabbrechern: Von den 83 randomisierten Teilnehmern wurden 4 von der Analyse ausgeschlossen, weil sie entweder nicht die Teilnahmebedingungen erfüllten oder vor der ersten Untersuchung die Studie abbrachen. 23 weitere Patienten brachen die Studie vor Ablauf der 6 Monate ab. Analysen wurden mit den Daten aller Studienteilnehmer (intent-to-treat, ITT) sowie nur mit den vollständigen Datensätzen (Completers) durchgeführt. Punkte: 1/1
Studienendpunkte	- Visual Analogue Scale für Schmerzen - Anzahl der geschwollenen Gelenke - Erythrozyten-Sedimentations-Rate - Arbeitszeitverluste - Symptomlosigkeit
Arbeitsbezogene Endpunkte	Arbeitszeitverluste, berechnet mit ITT-Population sowie nur mit Completers. Statistische Methoden sind den Ergebnissen nicht zu zuordnen.

Autor & Jahr	He et al. 2005
Erkrankung	Chronische Nackenschmerzen
Probanden	24 randomisiert Teilnahmebedingungen: Alter 20-50 Jahre; weibliches Geschlecht; beschäftigt in sitzender Position oder die Aufgabenfelder umfassten leichte, sich wiederholende Aufgaben; Nackenschmerzen, die im letzten Jahr mindestens 3 Monate anhielten und Arbeit und Allgemeinbefinden beeinträchtigten Ausschlusskriterien: Diabetes, Arbeitsbefreiung aufgrund der Nackenschmerzen; neuronale, rheumatische oder andere Erkrankungen; Schwangerschaft/Stillzeit; andere Schmerzbehandlung außer der Studienbehandlung Probandencharakteristika: Durchschnittsalter in Jahren: 47,3 Anteil weiblicher Probanden: 100 % durchschnittliche Jahre seit Diagnose: nicht berichtet
Behandlungen (Interventionen)	Alle Patienten erhielten 3 bis 4 Wochen lang, dreimal pro Woche eine Kombinationstherapie aus Akupunktur, Elektrostimulation und Ohr-Akupressur. Bei der Kontrollgruppe wurden absichtlich nicht empfohlene Stimulationspunkte gewählt.
Qualität (OQS)	Punkte: 2/5 Randomisierung: Die genaue Randomisierungsmethode wurde nicht beschrieben. Es wird jedoch erklärt, dass ein Zufallsmodell mit „Zurücklegen“ verwendet wurde. Punkte: 1/2 Verblindung: Die Behandlung wurde von einem der Studienautoren selbst durchgeführt. Eine Doppelverblindung ist daher nicht anzunehmen. Punkte: 0/2 Umgang mit Studienabbrechern: Es wurde berichtet, dass es keine Studienabbrecher gab. Punkte: 1/1

Studienendpunkte	- Schlafqualität - Einschränkungen am Arbeitsplatz - Einschränkungen in der Freizeit - Reizbarkeit und Schonverhalten - Müdigkeit - Frequenz des Schmerzes - Lebenszufriedenheit
Arbeitsbezogene Endpunkte	Die Einschränkungen am Arbeitsplatz wurden mit einem t-Test verglichen. Des weiteren wurde der Verlauf der Durchschnittswerte während der Studie und in den nachfolgenden Untersuchungsterminen beschrieben.

Autor & Jahr	Jarzem et al. 2005
Erkrankung	Chronische Rückenschmerzen
Probanden	350 randomisiert (324 analysiert) Teilnahmebedingungen: Alter zwischen 18 und 70 Jahren; anhaltende Rückenschmerzen seit mindestens 3 Monaten Ausschlusskriterien: Zentrum des Schmerzempfindens über T12; keine Schmerzausstrahlung in die unteren Extremitäten; vorhergehende Behandlung mit TENS, Kortikosteroiden, Antikoagulantien; gestellter Antrag auf Sozialleistungen bezüglich der Rückenschmerzen; vergangene oder aktive maligne Tumorerkrankung; Herzschrittmacherimplantat; schwere Allgemeinerkrankung; Schwangerschaft; operativer Eingriff während der letzten 3 Monate; während der Studie durchgeführte physiotherapeutische oder chiropraktische Behandlung Probandencharakteristika der analysierten Patienten: Durchschnittsalter in Jahren: 45,1 Anteil weiblicher Probanden: 50% durchschnittliche Jahre seit Diagnose: 10,2
Behandlungen (Interventionen)	Die Patienten erhielten eines von 4 TENS Geräten zur Eigentherapie für 3 Monate. 93 Patienten erhielten ein konventionelles TENS Gerät, 81 ein Akupunktur-TENS-Gerät, 86 ein NuWave-TENS-Gerät (biphasisch) und 89 Patienten bekamen ein funktionsloses Kontrollgerät. Die Geräte wurden durchschnittlich 188 Minuten am Tag getragen.
Qualität (OQS)	Punkte: 5/5 Randomisierung: Die Randomisierung wurde mit einer Zufallszahlentabelle ihren Behandlungsgruppen zugeteilt. Punkte: 2/2 Verblindung: Das Kontrollgerät war äußerlich nicht von einem funktionierenden zu unterscheiden. Es täuschte mit aktiven Lämpchen Funktion vor. Punkte: 2/2 Umgang mit Studienabbrechern: 350 Patienten wurden randomisiert. 1 Proband beendete die Studie bevor er ein Gerät erhalten hatte. 25 Patienten wurden aus der Studie entfernt, da sie nicht zu allen Untersuchungsterminen erschienen waren. Punkte: 1/1
Studienendpunkte	- McGill Activity Scale - McGill Work Scale - Roland Morris Disability Questionnaire (RMDQ-24) - Zung Depression Scale - funktionelle Messungen nach Waddell et al. (1992) - VAS für Schmerz (nicht berichtet)
Arbeitsbezogene Endpunkte	McGill Work Scale: eine multivariante Varianzanalyse wurde durchgeführt

Autor & Jahr	Kavanaugh et al. 2006 Methodendetails aus Antoni et al. 2005
Erkrankung	Psoriasisarthritis
Probanden	200 randomisiert

	Teilnahmebedingungen: Diagnose auf Psoriasisarthritis seit mindestens 6 Monaten; aktiver Gelenkbefall (mindestens 5 geschwollene und mindestens 5 druckempfindliche Gelenke sowie C-reaktiver-Proteinwert von mindestens 15 mg/l und/oder Morgensteifigkeit von mindestens 45 Minuten); nicht erfolgreiche vorhergehende Behandlung mit DMARDs oder NSARs; mindestens eine Hautläsion welche einen Durchmesser von wenigstens 2 cm hat Ausschlusskriterien: aktive oder durchgemachte Tuberkuloseinfektion; chronische oder akute Infektionskrankheit; malignes Geschehen; Herzinsuffizienz; vorhergehende TNF-α-Hemmer-Exposition; nachweisbarer Rheumafaktor, intraartikuläre Injektionen in den letzten 4 Wochen Probandencharakteristika: Durchschnittsalter in Jahren: 46,8 Anteil weiblicher Probanden: 39 % durchschnittliche Jahre seit Diagnose: 8,0
Behandlungen (Interventionen)	Die Probanden erhielten 22 Wochen lang entweder Infliximab- (5mg/kg) (n=100) oder Placebo-Infusionen (n=100). Die Infusionen wurden in den Wochen 0, 2, 6, 14 und 22 durchgeführt. Die letzte Kontrolluntersuchung erfolgte in Woche 24. Konnte bis Woche 16 keine ausreichende Reduktion der geschwollenen und druckempfindlichen Gelenke verzeichnet werden (<10%), so erhielten die Placebopatienten in den Wochen 16, 18 und 22 ebenfalls Infliximab. Bekamen die Patienten bereits mindestens 3 Monate vor Studienbeginn stabile Dosen von Methotrexat (max. 25 mg/Woche), so durften diese weiter eingenommen werden. Gleiches galt für orale Kortikosteroide (Äquivalentdosis 10 mg/Tag Prednison). Topische Kortikosteroide zur Behandlung der Psoriasis-Läsionen waren nur für das Gesicht und die Leistengegend gestattet.
Qualität (OQS)	Punkte: 3/5 Randomisierung: Es wurde nach Untersuchungszentrum und Methotrexat stratifiziert. Das Randomisierungsverfahren selbst wurde nicht näher beschrieben. Punkte: 1/2 Verblindung: Die gegebenen Medikamente wurden von einem nicht verblindeten Mitarbeiter vorbereitet. Das Personal, welches die Infusionen verabreichte waren jedoch verblindet, ebenso wie die Untersucher. Visuell waren die Pulver (aktiv und passiv) nicht zu unterscheiden. Um die Verblindung zu wahren, wurden auch Patienten mit aktiver Behandlung in Woche 16 bei mangelnder Verbesserung aus der Studie entfernt. Diese erhielten dann Placeboinfusionen in den Wochen 16 und 18. Punkte: 2/2 Umgang mit Studienabbrechern: Alle Berechnungen wurden nur bis zum Messpunkt in der 14 Woche durchgeführt. Es wurde weder über Studienabbrecher berichtet, noch über deren statistischen Einfluss. Punkte: 0/1
Studienendpunkte	- American College of Rheumatology 20% response (ACR20) - Short Form 36 Health Survey Questionnaire (SF-36) - Psoriasis Area and Severity Index (PASI) - Beschäftigungsstatus (arbeitsfähig, nicht arbeitsfähig, arbeitend, nicht arbeitend) - Visual Analogue Scale (VAS) für Produktivität
Arbeitsbezogene Endpunkte	- Beschäftigungsstatus und die Anzahl der Personen, die verlorene Arbeitszeit in den letzten 4 Tagen vor Studienende vorzuweisen hatten, wurden mit Fisher's exact Test verglichen. - Die Veränderung der VAS-Produktivität wurde in einem Box-Plot dargestellt. Median-Unterschiede wurden mithilfe eines T-Tests verglichen. - Die Veränderung der SF-36 Unterpunkte, die Produktivitätsverluste und Arbeitszeitverluste betreffen, wurden zwischen der aktiven und passiven Gruppe verglichen. Signifikanzberechnungen wurden mit einem Chi-Quadrat-Test einer Cochran–Mantel–Haenszel-Statistik angestellt. Des weiteren wurden diese Unterpunkte in Relation mit ACR20- und PASI75-Response verglichen.

Autor & Jahr	Kavanaugh et al. 2009a Detailbeschreibungen entnommen aus den „Rheumatoid Arthritis Prevention of Structural Damage“ Studien (Keystone et al. 2008, RAPID 1) & (Smolen et al. 2009, RAPID 2)
Erkrankung	Rheumatoide Arthritis
Probanden	1601 randomisiert (RAPID 1 n=982; RAPID 2 n=619) Teilnahmebedingungen: Alter ≥ 18 Jahre; bestätigte Diagnose nach den Kriterien des American College of Rheumatology (ACR) für mindestens 6 Monate, aber nicht länger als 15 Jahre; Krankheitsaktivität zu Beginn

	der Studie (Erythrozytensedimentationsrate ≥ 30 mm/h oder C-reaktives Protein Wert von mehr als 15 mg/l); vorhergehende Medikation mit MTX für mehr als 6 Monate, wobei eine regelmäßige Einnahme von mehr als 10 mg/Woche in den letzten 2 Wochen vor Studienbeginnobligatorisch war Ausschlusskriterien: andere entzündliche Arthriden sowie sekundäre nicht entzündliche Arthriden; vorgehende Behandlung mit einem biologisch basiertem Wirkstoff während der letzten 6 Monate (im Falle von Etanercept oder Anakinra nur 3 Monate); bekannte Allergien gegen biologisch basierte Medikamente; bekannte Wirkungslosigkeit von TNF-bindender Medikamente; durchgemachte oder aktuelle Tuberkulose Infektion Probandencharakteristika: Durchschnittsalter in Jahren: 52,0 Anteil weiblicher Probanden: 82,6 % durchschnittliche Jahre seit Diagnose: 6,1
Behandlungen (Interventionen)	Die RAPID-1-Studie wurde 52 Wochen, die RAPID-2-Studie 24 Wochen lang durchgeführt. In beiden Studien erhielten die Patienten der aktiven Gruppen in den Wochen 0, 2 und 4 je 400 mg Certolizumab. Ab Woche 6 erhielten sie alle 2 Wochen entweder 200 mg Certolizumab (n=639) oder 400 mg Certolizumab (n=636). Die Placebogruppe erhielt über die gesamte Studienzeit hinweg Placebo (n=326). Zusätzlich nahmen die Patienten die Dosis Methotrexat weiterhin zu sich, mit der sie bereits vor Studienbeginn behandelt wurden. Erreichten die Patienten keine Verbesserung nach dem ACR20 Kriterium (20% Verbesserung) in den Wochen 12 und 14, so wurden die Patienten in der 16. Woche aus der Studie entfernt.
Qualität (OQS)	Punkte: 3/5 Randomisierung: Methodik nicht beschrieben Punkte: 1/2 Verblindung: Die Methodik der Doppel-Verblindung wurde nicht beschrieben Punkte: 1/2 Umgang mit Studienabbrechern: Von den 982 der RAPID-1-Studie brachen 410 Patienten die Studie vorzeitig ab. In Woche 16 waren bereits 63 % der Placebogruppe wegen fehlender Wirkung aus der Studie ausgeschieden. In der RAPID-2-Studie brachen 247 Probanden vorzeitig ab. 79.5% der Placebo-Teilnehmer verließen die Studie in der 16. Woche. Für die Berechnungen wurden die letzten beobachteten Werte verwendet (LOCF). Punkte: 1
Studienendpunkte	Rheumatoid Arthritis Specific Work Productivity Survey (WPS-RA)
Arbeitsbezogene Endpunkte	- WPS-RA: Unterpunkte: > Abwesenheit (Arbeitsfehltage pro Monat) > Anwesenheit trotz Krankheit (Tage mit Produktivität > 50% pro Monat) > Interferenz mit Arbeitsproduktivität pro Monat Statistische Vergleiche der unterschiedlichen Behandlungen wurden mit einer nicht-parametrischen Bootstrap-Methode berechnet.

Autor & Jahr	Kavanaugh et al. 2013 Studiendetails aus Kavanaugh et al. 2009b
Erkrankung	Psoriasisarthritis
Probanden	405 randomisiert Teilnahmebedingungen: bestehende Psoriasisarthritis (≥ 3 geschwollende, ≥ 3 druckempfindliche Gelenke, Psoriasisläsion ≥ 2 cm im Durchmesser) trotz DMARD- oder NSAR-Behandlung Ausschlusskriterien: vorherige Behandlung mit TNF-Inhibitoren, Rituximab, Natalizumab oder mit zytotoxischen Medikamenten Probandencharakteristika: Durchschnittsalter in Jahren: 47,0 Anteil weiblicher Probanden: 39,8 % durchschnittliche Jahre seit Diagnose: 7,5
Behandlungen (Interventionen)	Die Patienten erhielten alle 4 Wochen subkutan entweder Golimumab 50 mg (n=146), Golimumab 100 mg (n=146) oder Placebo.

	Nahm die Anzahl der betroffenen Gelenke nicht um 10% ab, so erhielten ab der 16. Woche die Placebo-Patienten 50mg Golimumab, die 50mg-Golimumab-Patienten 100mg-Golimumab und die 100mg-Golimumab -Patienten weiterhin 100mg Golimumab. Ab Woche 24 wechselten alle Placebo-Patienten in die 50mg-Gruppe. Ab Woche 52 konnte die Dosis auf 100 mg Golimumab erhöht werden. Die Studie dauerte insgesamt 104 Wochen. Bereits bestehende, regelmäßige Einnahme von NSARs, Methotrexat oder Kortikosteroide (nicht mehr als 10 mg Prednison pro Tag) konnte während der Studie fortgesetzt werden.
Qualität (OQS)	Punkte: 4/5 Randomisierung: Die Randomisierung wurde von einem computergestützten Sprachdialogsystem durchgeführt. Stratifiziert wurde nach Behandlungszentrum und Methotrexat-Einnahme. Punkte: 2/2 Verblindung: Es wurde nicht darüber berichtet welche Maßnahmen für die Wahrung der Doppel-Blindheit getroffen wurden (ausgenommen das Sprachdialogsystem). Punkte: 1/2 Umgang mit Studienabbrechern: 25 Patienten beendeten die Studie bis in Woche 24. 22 weiter verließen die Studie bis Woche 52 und 23 weitere schieden bis Woche 104 aus. Analysen bezogen auf die Wochen 16 und 14 erfolgten mit der ITT-Herangehensweise. Fehlende Daten von Abbrechern und Patienten , die die Behandlung gewechselt hatten, wurden mit der LOCF-Methodik einbezogen. Punkte: 1/1
Studienendpunkte	- Health Assessment Questionnaire (HAQ) - Short Form 36 Health Survey Questionnaire (SF-36) - VAS für Produktivität - ≥75% improvement in Psoriasis Area and Severity Index (PASI75) - American College of Rheumatology 20% Response (ACR20) - 28-joint Count Disease Activity Score - C-Reactive Protein (DAS28-CRP) - PsA-modified Maastricht Ankylosing Spondylitis Enthesitis Score - Fingerentzündung
Arbeitsbezogene Endpunkte	VAS für Produktivität innerhalb der letzten 4 Wochen (0 keine Einschränkungen, 10 maximale Einschränkung) Für die Wochen 16 und 24 erfolgten ITT-Analysen mit dem Van der Waerden Varianz Test. Fehlende Daten wurden mit der LOCF-Methodik erfasst.

Autor & Jahr	Lehmann et al. 1986
Erkrankung	Chronische Rückenschmerzen
Probanden	54 randomisiert 53 analysiert (1 Patient lehnte die ihm zugewiesene Behandlung ab) Teilnahmebedingungen: Schmerzen in der Lumbalregion für mindestens 3 Monate; daraus resultierende Einschränkungen sollten den Behandlungsaufwand rechtfertigen Ausschlusskriterien: Proband benötigt spinalen Eingriff; Osteomyelitis; Tumorerkrankungen; Spondylitis ankylosans; Wirbelfrakturen; Skoliose; Schwangerschaft; Spondylodiszitis Probandencharakteristika der analysierten Patienten: Durchschnittsalter in Jahren: 40 Anteil weiblicher Probanden: 32,1 % Rückenschmerzen seit über 6 Monaten: 100%
Behandlungen (Interventionen)	Die Probanden wurden 3 Wochen lang entweder mit Transkutaner Elektrischer Nervenstimulation (TENS) (n=18), einem funktionslosen TENS-Gerät (n=18) oder mit Elektroakupunktur (n=17) behandelt. TENS Behandlungen wurden werktags täglich angeboten, Akupunktursitzungen zweimal in der Woche. Alle Gruppen erhielten eine ärztliche Beratung über Rückenschmerzen und nahmen zweimal täglich an sportlichen Übungen teil. Nach Ablauf der 3 Wochen sowie 6 Monate nach Entlassung wurden die Probanden erneut befragt.
Qualität (OQS)	Punkte: 4/5 Randomisierung: Es wurde nach Rückenoperationsstatus stratifiziert (OP während der letzten 6 Monate; OP liegt länger als 6, aber weniger als 12 Monate zurück; keine OP erfolgt, oder liegt länger als 12 Monate zurück). Über das Verfahren der Randomisierung wurde nicht berichtet.

	Punkte: 1/2 Verblindung: Nach Einteilung der Behandlungsgruppen wurden die TENS Geräte von einer unverblindeten Schwester eingestellt (unter die Wahrnehmungsgrenze der Patienten). Danach wurden die Batterien aus den Sham-Geräten (für die Scheinbehandlung) durch leere Batterien ersetzt. Patienten, Untersucher und Behandler waren verblindet. Punkte: 2/2 Umgang mit Studienabbrechern: 11 Patienten beendeten die Studie vorzeitig. Es wurde erwähnt, dass auch ihre Daten in die Analysen einflossen, jedoch wird über den genaueren statistischen Umgang mit den Abbrechern nichts berichtet. Punkte: 1/1
Studienendpunkte	- Bewegungsumfang - Stärke des Körperstammes - Spitzen- und Durchschnittschmerz (VAS 10cm) - von Patienten eingeschätzte Arbeitsfähigkeit - Rückkehr in ein Arbeitsverhältnis - Eingenommene Analgetika
Arbeitsbezogene Endpunkte	Rückkehr in ein Arbeitsverhältnis: Genauere, nach Behandlungsgruppen aufgeschlüsselte Berichte fehlen. Es wurde erwähnt, dass es keinen signifikanten Unterschied zwischen den Gruppen gab.

Autor & Jahr	Licciardone et al. 2003
Erkrankung	Chronische Rückenschmerzen
Probanden	91 randomisiert Teilnahmebedingungen: Alter 21-69 Jahre; Schmerzen in der Lumbalregion für mindestens 3 Monate Ausschlusskriterien: Osteomyelitis oder Fraktur in der lumbalen Region; Bandscheibenvorfall; Spondylitis ankylosans; Cauda-equina-Kompressions-Syndrom; vergangene oder aktive maligne Erkrankung; Rückenoperation oder manuelle Therapie während der letzten 3 Monate; Schwangerschaft/Stillzeit; Erhalt von Sozialleistungen während der letzten 3 Monate erhalten; Mitarbeiter im Behandlungszentrum während der letzten 3 Monate Probandencharakteristika: Durchschnittsalter in Jahren: 49,8 Anteil weiblicher Probanden: 65,1 % Rückenschmerzen seit über einem Jahr: 51,1 %
Behandlungen (Interventionen)	Die Patienten erhielten für die Dauer von 5 Monaten monatlich entweder eine osteopatische (n=48) Intervention, eine osteopathische Scheinbehandlung (n=23) oder gar keine Behandlung (n=20).
Qualität (OQS)	Punkte: 3/5 Randomisierung: Die Gruppenzuteilung wurde durch das Ziehen von versiegelten Briefumschlägen realisiert. Punkte: 2/2 Verblindung: Das durchführende Personal war nicht verblindet. Verblindung bestand nur für die restlichen Forscher (Untersucher, Interviewpersonal) und Patienten. Punkte: 0/2 Umgang mit Studienabbrechern: 25 Probanden beendeten die Studie vorzeitig. Wie mit den Abbrechern statistisch verfahren wurde ist nicht beschrieben. Punkte: 1/1
Studienendpunkte	- Short Form 36 Health Survey Questionnaire (SF-36) - VAS für Schmerzen - Roland Morris Disability Questionnaire (RMDQ-24) - verlorene Arbeits- oder Schultage - Zufriedenheit mit Behandlung
Arbeitsbezogene Endpunkte	verlorene Arbeits- oder Schultage in den letzten 4 Wochen: Da keine signifikanten Unterschiede zwischen den Behandlungsgruppen gefunden werden konnte, wurde auf eine genauere Beschreibung der Ergebnisse verzichtet.

Autor & Jahr	Manchikanti et al. 2010
Erkrankung	Chronische Nackenschmerzen
Probanden	120 randomisiert Teilnahmebedingungen: Alter ≥ 18 Jahre; einschränkende Nackenschmerzen seit mindestens 6 Monaten; positives Ansprechen auf temporäre Ramus medialis Blockade (80% Schmerzverbesserung und verringerte Bewegungseinschränkung) Ausschlusskriterien: Bandscheibenvorfall; spinale Stenose; spinaler Eingriff im zervikalen Bereich während der letzten 3 Monate; instabile psychische Erkrankungen; Einnahme von mehr als 300 mg Opioid (äquivalent zu Morphin); schwere Allgemeinerkrankung; andere Schmerzerkrankungen; Schwangerschaft/Stillzeit; bekannte Unverträglichkeiten gegen Lokalanästhetika, Sarapin oder Steroide; Patienten, die für die Behandlung nicht ordnungsgemäß positioniert werden können Probandencharakteristika: Durchschnittsalter in Jahren: 44,5 Anteil weiblicher Probanden: 74,2 % durchschnittliche Jahre seit Diagnose: 8,6
Behandlungen (Interventionen)	Alle Patienten wurden mit diagnostischen Nervenblockaden auf das Ansprechen der Behandlung getestet. Nach Randomisierung erhielt Gruppe I Injektionen mit Bupivacain (n=60) und Gruppe II (n=60) Injektionen mit einer Bupivacain-Steroid-Mischung. Zusätzlich wurden beide Behandlungsgruppen in 2 Untergruppen unterteilt. Bei je einer Gruppe enthielt das Injektionsgemisch zusätzlich Sarapin. Die Patienten wurden 2 Jahre lang alle 3 Monate behandelt und untersucht. Reagierten die Patienten nicht mit fallenden Schmerzwerten (mind. 50%) so wurden sie aus der Studie entfernt.
Qualität (OQS)	Punkte: 5/5 Randomisierung: Es wurde eine computergestützte Blockrandomisierug durchgeführt. Punkte: 2/2 Verblindung: Die OP-Schwestern bereiteten die Injektionen vor. Patienten und Behandler waren verblindet. Alle vorzeitig entblindeten Patienten wurden aus der Studie entfernt. Die Injektionsgemische waren visuell nicht voneinander zu unterscheiden. Punkte: 2/2 Umgang mit Studienabbrechern: 26 Patienten beendeten die Studie vorzeitig. Alle statistischen Berechnungen wurden mit der ITT- und LOCF-Herangehensweise durchgeführt. Punkte: 1/1
Studienendpunkte	- Nackenschmerzen auf einer Skala von 0 bis 10 - Neck Disability Index (NDI) - Opioideinahme - Beschäftigungsstatus
Arbeitsbezogene Endpunkte	Beschäftigungsstatus und Arbeitsfähigkeit: Der Status der Patienten zu drei Zeitpunkten (3, 12, 24 Monate) wurde in tabellarischer Form berichtet.

Autor & Jahr	Markenson et al. 2005
Erkrankung	Osteoarthritis
Probanden	109 randomisiert (107 erhielten Behandlung) Teilnahmebedingungen: bestätigte Osteoarthritis-Diagnose nach ACR; andauernde Schmerzen seit einem Monat oder nach Absetzen von Opioiden; Schmerzen in der Woche vor Studienbeginn von mindestens 5/10 (mind. 3 wenn Opioide eingenommen wurden); seit mindestens 2 Wochen folgende Medikamentation: NSAR- oder Paracetamol-Einnahme oder orale Opioide (weniger als 60 mg/Tag äquivalent zu Oxycodon) mit oder ohne zusätzliche NASR/Paracetamol-Gabe Ausschlusskriterien: bekannte Allergie gegen Opioide; geplanter Eingriff während Studie; bestehende schwere Allgemeinerkrankung; maligne Tumorerkrankung; Schwangerschaft/Stillzeit; bestehende oder vergangene Drogenprobleme; bestehende Rechtsstreitigkeiten bezüglich der Erkrankung; intraartikuläre oder intramuskuläre Injektionen mit Steroiden während der vergangenen 6 Wochen. Probandencharakteristika: Durchschnittsalter in Jahren: 63 Anteil weiblicher Probanden: 72,9 %

	durchschnittliche Jahre seit Diagnose: nicht berichtet
Behandlungen (Interventionen)	Die Probanden nahmen 3 Monate lang alle 12 Stunden entweder 10 mg Oxicodon (n = 56) oder Placebo (n = 51) zu sich. Die Dosis konnte bis auf 120 mg/Tag erhöht werden. Es wurde versucht eine stabile Tagesdosis zu finden, definiert als die Dosis, mit der ein Schmerzwert von maximal 4/10 über 48 Stunden erreicht werden konnte. Die bereits vor Studieneintritt gegebenen Dosen an NSAR/Paracetamol durften weiterhin eingenommen werden.
Qualität (OQS)	Punkte: 4/5 Randomisierung: Die Randomisierungssequenz wurde computergestützt durchgeführt. Punkte: 2/2 Verblindung: Die Studien-Medikamente wurden in mit Zufallscodes versehenen Flaschen vom Sponsor geliefert. Über andere Eigenschaften der Tabletten wurde nicht berichtet. Punkte: 1/2 Umgang mit Studienabbrechern: 109 Patienten wurden randomisiert, 2 Patienten beendeten die Studie bevor die ersten Medikamente ausgegeben wurden. Von den 107 analysierten Patienten beendeten 71 die Studie vorzeitig. Alle Endpunkte wurden mit der LOCF-Methodik berechnet. Punkte: 1/1
Studienendpunkte	Primäre Endpunkte: - Brief Pain Inventory (BPI) Durchschnittsschmerz bei stabiler Dosis - Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) an den Tagen 30 und 60 - Anzahl der Studienabbrecher, die wegen fehlender Wirkung die Studie verließen Sekundäre Endpunkte: - BPI Unterpunkte: Schmerz, Interferenz, Funktion - WOMAC (alle Untersuchungstage) - Patient's Global Impression (PGI) - Tage bis stabile Dosis gefunden wurde - Dosishöhe - Patientenzufriedenheit bezogen auf Behandlung
Arbeitsbezogene Endpunkte	BPI: Unterpunkt Interferenz mit normaler Arbeit; Eine Kovarianzanalyse wurde durchgeführt, Kovarianten: Baslinewerte, Altersgruppe, Geschlecht. Fehlende Daten wurden mit LOCF einbezogen.

Autor & Jahr	Meireles et al. 2010
Erkrankung	Rheumatoide Arthritis
Probanden	82 randomisiert Teilnahmebedingungen: rheumatoide Arthritis mit mindestens 2 entzündeten Gelenken der Hand (Handgelenke, proximale oder mediale Phalanxgelenke); Schmerzen während alltäglicher Beschäftigungen (VAS 3-8/10) Ausschlusskriterien: Hautveränderungen am geplanten Behandlungsbereich; andere gesundheitliche Beeinträchtigungen der Hand; Physiotherapie oder Gelenkinjektionen während der vorhergehenden 3 Monate; Veränderungen der Medikationsparameter während der letzten 3 Monate (Kortikosteroide und DMARDs) Probandencharakteristika: Durchschnittsalter in Jahren: 52,8 Anteil weiblicher Probanden: 97,6 % durchschnittliche Jahre seit Diagnose: 11,8
Behandlungen (Interventionen)	Über die Dauer von 2 Monaten erhielt die aktive Gruppe (n=41) 2mal wöchentlich eine Behandlung mit einem Niedrig-Energie-Laser (3 J/cm^2 bei 785 nm), während die Kontrollgruppe (n=41) entsprechend mit einem Placebogerät ohne Emission behandelt wurde.
Qualität (OQS)	Punkte: 5/5 Randomisierung: durchgeführt mit einer computergenerierten (Excel) Zufallszuordnung. Punkte: 2/2 Verblindung: Die Gruppenzuordnung wurde in undurchsichtigen, versiegelten Briefumschlägen festgehalten. Während der Behandlung wurden Hand und Gerät mit einem Tuch überdeckt. Weder Patienten, Forscher

	oder durchführende Therapeuten wussten, welche Therapie durchgeführt wurde. Punkte: 2/2 Umgang mit Studienabbrechern: 4 Patienten beendeten die Studie vorzeitig. Alle befanden sich in der Placebogruppe und verließen die Studie aus persönlichen Gründen, ausgenommen ein Patient, der die Behandlung als unangenehm empfand. Die statistischen Berechnungen wurden mit den Daten aller randomisierten Patienten durchgeführt. Es wurde jedoch nicht berichtet, ob die LOCF oder die BOCF Methodik verwendet wurde. Punkte: 1/1
Studienendpunkte	Primäre Endpunkte: - VAS-Schmerzskala während alltäglichen Aktivitäten(0 cm–10 cm), generell und für jedes Gelenk - Health Assessment Questionnaire (HAQ) - Disabilities of the Arm, Shoulder and Hand Questionnaire (DASH) Sekundäre Endpunkte: - Likert Skala (1-5, 1 keine Entzündung, 5 sehr starke Entzündung) - Morgensteifigkeit (Intensität untersucht mit VAS, sowie Dauer in Minuten) - Greifkraft der Finger und der gesamten Hand - O'Connor Test für Beweglichkeit - Bewegungsumpfang der einzelnen Gelenke der Hand - Umfang der Gelenke - Likert Skala für Patientenzufriedenheit
Arbeitsbezogene Endpunkte	Disabilities of the Arm, Shoulder and Hand Questionnaire (DASH), Unterpunkt: „work domain“ Eine Varianzanalyse wurde durchgeführt.

Autor & Jahr	Skljarevski et al. 2009
Erkrankung	Chronische Rückenschmerzen
Probanden	404 randomisiert Teilnahmebedingungen: Schmerzen des unteren Rückens (Wirbel T6 und abwärts) an den meisten Tagen der vergangenen 6 Monate oder länger; wöchentlicher Durchschnittsschmerzwert ≥ 4/10; Ausstrahlung der Schmerzen nur bis in das proximale Glied der unteren Extremitäten (QTFC I & II) Ausschlusskriterien: Wurzelkompressionssyndrom; Stenose des Spinalkanales; Spondylolisthesis des Schweregrades III und IV; mehr als eine oder eine Rückenoperation während des letzten Jahres; BMI > 40 kg/m^2; vorliegen einer schweren depressiven Episode; invasive Rückenbehandlungen im vorangegangenen Monat; Bewerbung auf den Erhalt von finanziellen Zuwendungen aufgrund von Arbeitsunfähigkeit durch Rückenschmerzen Probandencharakteristika: Durchschnittsalter in Jahren: 53,9 Anteil weiblicher Probanden: 57,4 % durchschnittliche Jahre seit Diagnose: 11,7
Behandlungen (Interventionen)	Die Patienten erhielten 13 Wochen lang täglich entweder Placebo(n=117) oder Duloxetin in den Dosen 20 mg (n=59), 60 mg (n=116) oder 120 mg (n=112). Wurden die Patienten bereits vor der Studie regelmäßig mit NSARs oder Physiotherapie therapiert, so durfte dies unter Beibehaltung der Dosis bzw. der Frequenz der Behandlung fortgesetzt werden. Es war den Probanden erlaubt, Schmerzmittel mit kurzer Wirkung gegen akute Leiden oder als Rettungsmedikation zu sich zu nehmen. Die Verwendung wurde auf maximal 3 aufeinander folgende Tage und nicht mehr als insgesamt 20 Tage beschränkt.
Qualität (OQS)	Punkte: 5/5 Randomisierung: Die Patienten wurden mithilfe einer computergenerierten Zufallssequenz, stratifiziert nach NSARs-Einnahme, ihren Behandlungsgruppen zugeteilt. Punkte: 2/2 Verblindung: Um die Verblindung von Patienten, Behandlern und Forschern zu wahren, wurde bei der Gruppenzuteilung ein Sprachdialogsystem verwendet. Die Kapseln zur Einnahme waren allesamt identisch in Aussehen, Geschmack und Geruch. Punkte: 2/2 Umgang mit Studienabbrechern: 137 Probanden beendeten die Studie vorzeitig. Nur in der 120 mg/Tag

	Duloxetin-Gruppe beendeten signifikant mehr Patienten wegen Nebenwirkungen die Studie als in der Placebo-Gruppe. Ansonsten gab es zwischen den Gruppen keinen signifikanten Unterschied bezüglich des Abbruchgrundes. Eine MMRM-Analyse wurde verwendet um den primären Endpunkt darzustellen. Sekundäre Endpunkte wurden mit der LOCF-Methodik in einer Kovarianzanalyse untersucht. Punkte: 1/1
Studienendpunkte	Primäre Endpunkte: - Durchschnittliche Veränderung (Baseline-Endpunkt) des wöchentlichen 24h Durchschnittsschmerzes bei 60 mg/Tag gegenüber Placebo Sekundäre Endpunkte: - Patient's Global Impressions of Improvement (PGI-I) - Roland-Morris Disability Questionnaire-24 (RMDQ-24) - BPI-S, BPI-I - Clinical Global Impressions-Severity of Illness (CGI-S) - Ansprechen der Therapie (30%, 50% Verbesserung des Schmerzes gegenüber Baseline) - Athens Insomnia Scale - Short Form 36 Health Survey Questionnaire (SF-36) - EuroQoL Questionnaire: 5 Dimensions (EQ-5D) - Beck Depression Inventory-II (BDI-II) (nicht berichtet) - Hospital Anxiety Depression Scale anxiety subscale (HADS-A) (nicht berichtet)
Arbeitsbezogene Endpunkte	BPI-I: Unterpunkt Interferenz mit normaler Arbeit ANCOVA mit Baselinewerten als Kovariante

Autor & Jahr	Skljarevski et al. 2010a
Erkrankung	Chronische Rückenschmerzen
Probanden	236 randomisiert Teilnahmebedingungen: Rückenschmerzen an den meisten Tagen der vergangenen 6 Monate im unteren Rückenbereich (Wirbel T6 und abwärts); 24h-Durchschnittsschmerzwert ≥ 4 (Schmerztagebuch, 0 keine Schmerzen, 10 schlimmst vorstellbare Schmerzen); Ausstrahlung der Schmerzen nur bis in das proximale Glied der unteren Extremitäten (QTFC I & II) Ausschlusskriterien: Wurzelkompressionssyndrom; Stenose des Spinalkanales; Tumorerkrankungen oder Abszesse in der Umgebung des Rückens; Spondylolisthesis des Schweregrades III und VI; Schwangerschaft/Stillzeit; mehr als eine oder eine Rückenoperation während des letzten Jahres; vorhergehende Teilnahme an einer Duloxetin-Studie; schwere psychische Erkrankungen; BMI > 40 kg/m^2; Behandlung mit Monoaminooxidase-Hemmern während der letzten 2 Wochen vor Randomisierung; weniger als 70% des Schmerztagebuches während der Screeningphase ausgefüllt. Probandencharakteristika: Durchschnittsalter in Jahren: 52 Anteil weiblicher Probanden: 61 % durchschnittliche Jahre seit Diagnose: 9,2
Behandlungen (Interventionen)	Die Patienten erhielten 13 Wochen lang täglich entweder 60 mg Duloxetin (n=115) oder Placebo (n=121). Hatte sich der 24h-Durchschnittsschmerz (BPI) nicht um mindestens 30% verbessert, so wurde die Dosis in Woche 7 für die aktive Gruppe auf 120 mg Duloxetin erhöht. Erhielten die Patienten bereits vor der Studie regelmäßig NSARs oder Paracetamol, so durften diese unter Beibehaltung der Dosis weiter eingenommen werden. Es war den Probanden erlaubt episodisch NSARs gegen akute Leiden zu verwenden. Die Verwendung wurde auf maximal 3 aufeinander folgende Tage und nicht mehr als insgesamt 20 Tage beschränkt.
Qualität (OQS)	Punkte: 4/5 Randomisierung: Es wurde eine stratifizierte Randomisierung nach NSAR Einnahme durchgeführt. Eine weitere Beschreibung des Verfahrens erfolgte nicht. Punkte: 1/2 Verblindung: Eine entsprechende, scheinbare Dosiseskalation in Woche 7 wurde ebenfalls für die Placeboanwender durchgeführt. Punkte: 2/2 Umgang mit Studienabbrechern: 54 Probanden beendeten die Studie vorzeitig. Signifikant mehr

	Studienabbrecher wegen Nebenwirkungen wurde in der aktiven Behandlungsgruppe festgestellt. Der 24h-Durchschnittsschmerz wurde in einem MMRM-Modell dargestellt. Die ANCOVA-Modelle verwendeten die LOCF-Methodik. Punkte: 1/1
Studienendpunkte	Primäre Endpunkte: - BPI (24h-Durchschnittsschmerz, BPI-S, BPI-I) Sekundäre Endpunkte: - Patient's Global Impressions of Improvement (PGI-I) - Roland-Morris Disability Questionnaire-24 (RMDQ-24) - BPI-S, BPI-I - Clinical Global Impressions-Severity of Illness (CGI-S) - Athens Insomnia Scale - Short Form 36 Health Survey Questionnaire (SF-36) - EuroQoL Questionnaire: 5 Dimensions (EQ-5D) - Work Productivity and Activity Impairment instrument (WPAI) - Beck Depression Inventory-II (BDI-II) - Hospital Anxiety Depression Scale anxiety subscale (HADS-A)
Arbeitsbezogene Endpunkte	- BPI-I: Unterpunkt Interferenz mit normalem Arbeitsablauf - WPAI: Unterpunkt Aktivitätseinschränkung am Arbeitsplatz, andere Unterpunkte nicht genauer berichtet Es wurden Kovarianzanalysen durchgeführt mit Baselinewerten und LOCF Endpunkten, stratifiziert nach NSAR-Einnahme.

Autor & Jahr	Skljarevski et al. 2010b
Erkrankung	Chronische Rückenschmerzen
Probanden	401 randomisiert Teilnahmebedingungen: Schmerzen des unteren Rückens an über der Hälfte der Tage während der letzten 6 Monate; Durchschnittsschmerzwert (BPI) ≥ 4; Ausstrahlung der Schmerzen nur bis in das proximale Glied der unteren Extremitäten (QTFC I & II) Ausschlusskriterien: benötigte tägliche Medikation mit NSARs, Paracetamol oder Opioiden; Schwangerschaft/Stillzeit; mehr als eine Rückenoperation oder eine Rückenoperation während des letzten Jahres; vorhergehende Teilnahme an einer Duloxetin-Studie; schwere psychische Erkrankungen; BMI > 40 kg/m^2; Behandlung mit Monoaminooxidase-Hemmern während der letzten 2 Wochen vor Randomisierung; Wurzelkompressionssyndrom; Stenose des Spinalkanales; Tumorerkrankungen oder Abszesse in der Umgebung des Rückens; Spondylolisthesis des Schweregrades III und VI Probandencharakteristika: Durchschnittsalter in Jahren: 54,1 Anteil weiblicher Probanden: 61,3 % durchschnittliche Jahre seit Diagnose: 8,5
Behandlungen (Interventionen)	Die Patienten erhielten 12 Wochen lang einmal täglich 60 mg Duloxetin (n=198) oder Placebo (n=203). Es war den Probanden erlaubt Notfallmedikation in Form von Ibuprofen, Paracetamol oder Naproxen zu verwenden. Die Verwendung wurde auf maximal 3 aufeinander folgende Tage und nicht mehr als insgesamt 20 Tage beschränkt.
Qualität (OQS)	Punkte: 3/5 Randomisierung: Die Art der Randomisierung wurde nicht beschrieben. Punkte: 1/2 Verblindung: Die verwendeten Verfahren der Doppel-Verblindung wurde nicht beschrieben. Punkte: 1/2 Umgang mit Studienabbrechern: 98 Patienten beendeten die Studie vorzeitig. In der Placebogruppe schieden signifikant mehr Probanden wegen mangelnder Wirkung aus, während in der Behandlungsgruppe signifikant mehr Probanden wegen Nebenwirkungen die Studie abbrachen. Der primäre Endpunkt wurden mit einer MMRM-Analyse berichtet. Die Baseline-Endpunkt-Kovarianzanalysen wurden mit der BOCF sowie der LOCF Methodik durchgeführt. Zusätzlich wurde eine modifizierte Version des BOCF verwendet, bei der nur Patienten, die als Studienaustrittsgrund Nebenwirkungen nannten, mit ihren Baseline-Daten eingeflossen sind, während alle anderen Probanden mit der LOCF Methode untersucht wurden. Die sekundären Endpunkte wurden mit der LOCF Methode errechnet.

	Punkte: 1/1
Studienendpunkte	Primäre Endpunkte: - BPI (Durchschnittliche Schmerzstärke der letzten 24h) Sekundäre Endpunkte: - Patient's Global Impressions of Improvement (PGI-I) - Roland Morris Disability Questionnaire (RMDQ-24) - BPI- Severity (BPI-S) - BPI-Interference (BPI-I) - Clinical Global Impressions-Severity (CGI-S) - Profile of Mood States-Brief Form (POMS-Brief Form) - Work Productivity and Activity Impairment (WPAI) - Short Form 36 Health Survey Questionnaire (SF-36) - European Quality of Life Questionnaire–5 Dimension (EQ-5D)
Arbeitsbezogene Endpunkte	- BPI-I: Unterpunkt Interferenz mit normalem Arbeitsablauf - WPAI: Unterpunkte: Abwesenheit, Anwesenheit mit Produktivitätseinbußen, Produktivitätsverlust und Aktivitätseinschränkung am Arbeitsplatz Die Kovarianzanalysen wurden mit Baseline adjustierten LOCF-Endpunktwerten durchgeführt.

Autor & Jahr	Smolen et al. 2006 Methodendetails aus St. Clair et al. 2004
Erkrankung	Rheumatoide Arthritis
Probanden	1049 randomisiert (1004 in dieser Publikation untersucht, Patienten im Alter von über 64 Jahren wurden nicht analysiert) Teilnahmebedingungen: Alter 18-75 Jahre; bestätigte Diagnose nach nach ACR; Synovisitis seit mintestens 3 Monaten aber nicht länger als 3 Jahre; mindestens 10 geschwollene und mindestens 12 druckempfindliche Gelenke; mindestens einer der folgenden Befunde: positiver Rheumafaktor; radiologische Erosionen an Hand- oder Fußgelenken; C-reaktives-Protein-Wert höher als 2,0 mg/ml. Ausschlusskriterien: vorherige Methotrexat-Exposition; Behandlung mit einem anderen DMARD in den letzten 4 Wochen oder Behandlung mit Leflunomid während der letzten 6 Monate; vorgehende Behandlung mit TNFα-Inhibitoren; bestehende Infektionserkrankung; durchgemachte maligne Tumorerkrankung in den letzten 5 Jahren Probandencharakteristika (randomisiert): Durchschnittsalter in Jahren: 50,4 Anteil weiblicher Probanden: 71 % durchschnittliche Jahre seit Diagnose: 0,9
Behandlungen (Interventionen)	Die Probanden erhielten in den Wochen 0, 2, 6 und fortlaufend jede 8. Woche Infusionen mit Infliximab 3 mg/kg (n = 359), 6 mg/kg (n = 363) oder Placebo (n = 282). Zusätzlich wurde Methotrexat (bis zu 20 mg/Woche) und Folsäure (5 mg/Woche) gegeben. Die Studie dauerte 54 Wochen. Erhielten die Patienten vor Studienbeginn orale Kortikosteroide, so durften diese unter Beibehaltung einer stabilen Dosis weiter eingenommen werden (maximal 10 mg/Tag).
Qualität (OQS)	Punkte: 5/5 Randomisierung: Randomisiert wurde mit einem Computersystem. Stratifiziert wurde nach Behandlungszentrum. Punkte: 2/2 Verblindung: Bei der Zuteilung der Medikationen wurde ein Sprachdialogsystem verwendet. Vor Entblindung wurden 45 Patienten von der Analyse ausgeschlossen, weil die originalen Datenbestände nicht überprüft werden konnten. Punkte: 2/2 Umgang mit Studienabbrechern: 201 Probanden beendeten die Studie vorzeitig. Fehlende Daten flossen als LOCF in die Berechnungen ein. Punkte: 1/1
Studienendpunkte	- Arbeitsstatus (beschäftigungsfähig, beschäftigt, in Vollzeit beschäftigt) - Arbeitsfehlzeiten während der Studie (in Tagen)
Arbeitsbezogene	- Die Veränderung der Arbeitsfähigkeit (Baseline-Endpunkt) wurde mit dem McNemars Test untersucht.

Endpunkte	Es wurde der Odds-Ratio für die Verbesserung des Arbeitsstatus statt Verschlechterung berechnet. Die Assoziation von Beschäftigungsstatus und HAQ-Werten wurde mit dem Cochran–Armitage Test berichtet. - Unterschiede in Arbeitsfehlzeiten wurden mit der Chi-Quadrat-Statistik analysiert.

Autor & Jahr	Strand et al. 1999
Erkrankung	Rheumatoide Arthritis
Probanden	482 randomisiert (438 analysiert) Teilnahmebedingungen: Alter ≥ 18 Jahre; bestätigte Diagnose auf rheumatoide Arthritis nach den Kriterien des ACR seit mindestens 6 Monaten; bestehende Krankheitsaktivität definiert als Vorliegen von mindestens 3 der folgenden Befunde: ≥ 9 druckempfindliche Gelenke, ≥ 6 geschwollene Gelenke, Morgensteifigkeit für mindestens 45 Minuten, Erythrozytensedimentationsrate von mehr als 28 mm/h Ausschlusskriterien: vorherige Behandlung mit Methotrexat; Behandlung mit anderen DMARDs während des letzten Monats Probandencharakteristika: Durchschnittsalter in Jahren: 54 Anteil weiblicher Probanden: 73 % durchschnittliche Jahre seit Diagnose: 6,8
Behandlungen (Interventionen)	Die Patienten erhielten 12 Monate lang 20 mg/Tag Leflunomid (n=182) oder 7,5 - 15,0 mg/Woche Methotrexat (n=182) oder Placebo (n=118). Wurden die Patienten bereits vor der Studie regelmäßig mit NSARs und/oder mit Prednison (Dosis ≤ 10 mg/Tag) therapiert, so durfte dies unter Beibehaltung der Dosis fortgesetzt werden.
Qualität (OQS)	Punkte: 4/5 Randomisierung: Das Randomisierungsverfahren wurde nicht beschrieben. Punkte: 1/2 Verblindung: Um die Doppel-Verblindung zu wahren, erhielten alle Patienten einmal pro Tag Leflunomid/Placebo und einmal wöchentlich Methotrexat/Placebo. Punkte: 2/2 Umgang mit Studienabbrechern: Patienten wurden in die Analyse mit einbezogen, wenn sie mindestens einmal nach Baseline die HAQ und SF-36 Fragebögen ausgefüllt hatten und die Kriterien des „Response Consitency Index“ erfüllten. So wurden von den 482 randomisierten Patienten nur 438 untersucht. Weitere fehlende Daten wurden mit der LOCF-Methodik einbezogen. Punkte: 1/1
Studienendpunkte	- Short Form 36 Health Survey Questionnaire (SF-36) - Health Assessment Questionnaire (HAQ) - Modified Health Assessment Questionnaire (MHAQ) - Die Top 5 des Problem Elicitation Technique (PET) (untersucht welche Unterpunkte des HAQ für die Patienten besonders wichtig sind und von der Krankheit eingeschränkt werden) - American College of Rheumatology Response Criteria 20%/50% - Arbeitsproduktivität
Arbeitsbezogene Endpunkte	Arbeitsproduktivität wurde mit einer modifizierten Frage aus der „Work Limitations Questionnaire of the National Opinion Research Center Survey“ untersucht. Die Varianzanalyse wurde gegen Placebo durchgeführt. Die fehlenden Daten wurden mit LOCF-Methodik einbezogen.

Autor & Jahr	Straube et al. 2011
Erkrankung	Fibromyalgie
Probanden	Die Daten von insgesamt 2757 Patienten, untersucht in 4 Studien, wurden in einer Meta-Analyse kombiniert. Teilnahmebedingungen: Alter ≥ 18 Jahre; bestätigte Diagnose nach den ACR Kriterien (1990); Schmerzen ≥ 40/100 mm auf VAS (ohne Schmerz- oder Schlafmedikamente)

	Probandencharakteristika der analysierten Patienten: Durchschnittsalter in Jahren: 49,0 Anteil weiblicher Probanden: 93,1 % durchschnittliche Jahre seit Diagnose: 9,1
Behandlungen (Interventionen)	In den 4 untersuchten Studien wurden Patienten mit Pregabalin (150 mg, 300 mg, 450 mg, 600 mg/Tag) oder Placebo behandelt. Zwei Studien wurden über 14 Wochen lang durchgeführt, die anderen beiden erstreckten sich über 8 und 13 Wochen.
Qualität (OQS)	Punkte: 5/5 Die Autoren führten selbst eine Bewertung der zu Grunde liegenden Studienberichte durch und vergaben für alle Studien 5 Punkte auf der OQS.
Studienendpunkte	- Fibromyalgia Impact Questionnaire (FIQ) - Short Form 36 Health Survey Questionnaire (SF-36) - Sheehan Disability Scale (SDS) - Multidimensional Assessment of Fatigue (MAF) - Visual Analog Scale (VAS)
Arbeitsbezogene Endpunkte	- FIQ Unterpunkte: Interferenz mit Arbeit, Arbeitszeitverluste - SF-36 Unterpunkt: Interferenz mit Arbeit - SDS Unterpunkt: Interferenz mit Arbeit - MAF Unterpunkt: Interferenz mit Arbeit Die Ergebnisse wurden aufgeschlüsselt nach Medikamentendosis, Schmerzverbesserung und Schmerzstatus. Signifikanztests wurden innerhalb der Behandlungsgruppen (Baseline–Endpunkt) mittels eines t-Tests durchgeführt.

Autor & Jahr	van der Heijde et al. 2006 Methoden Details aus van der Heijde et al. (2005)
Erkrankung	Spondylitis ankylosans
Probanden	279 randomisiert Teilnahmebedingungen: seit mindestens 3 Monaten bestätigte Diagnose von AS nach den modifizierten New York Kriterien; einen Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) von mindestens 4/10; VAS für Rückenschmerzen von mindestens 4/10; normaler Röntgen-Thorax Ausschlusskriterien: aktive oder durchgemachte Tuberkuloseinfektion; komplette Ankylosierung des Rückgrates; andere entzündliche, rheumatische Erkrankung; Fibromyalgie; Transplantatempfänger; Infektionserkrankung; maligne Tumorerkrankung; Multiple Sklerose; Herzinsuffizienz Probandencharakteristika: Durchschnittsalter in Jahren: 40,3 Anteil weiblicher Probanden: 19,3 % durchschnittliche Jahre seit Diagnose: 9,2
Behandlungen (Interventionen)	Die Probanden erhielten 5 Infusionen (jeweils in Woche 0, 2, 6, 12, 18) mit 5 mg/kg Infliximab (n=201) oder mit Placebo (n=78). Die Studie lief über 24 Wochen. Es durften zusätzlich stetige Dosen von NSARs, Paracetamol oder Tramadol gegeben werden.
Qualität (OQS)	Punkte: 4/5 Randomisierung: Das Verfahren der Randomisierung wurde nicht beschrieben. Es wurde nach Behandlungszentrum und C-reaktivem Protein-Wert stratifiziert. Punkte: 1/2 Verblindung: Beide Studienmedikamente waren visuell identisch (weißes Pulver). Punkte: 2/2 Umgang mit Studienabbrechern: 6 Patienten beendeten die Studie vorzeitig. Der Umgang mit den fehlenden Daten für die Berechnung der Statistik wurde nicht beschrieben. Punkte: 1/1
Studienendpunkte	- Short Form 36 Health Survey Questionnaire (SF-36): Unterpunkt Einfluss auf Arbeit und alltägliches - VAS für Arbeitsproduktivität (1-10) - Verlorene Arbeitszeit (in Tagen) - Korrelation von der Arbeitsproduktivität und des BASFI

	- Korrelation von der Verbesserung der Arbeitsproduktivität und der Verbesserung des BASFI
Arbeitsbezogene Endpunkte	- SF-36: McNemar-Test für Veränderung gegenüber Baseline - Produktivität (VAS): Die Produktivitätsverluste wurden zwischen der aktiven und der Placebogruppe verglichen. Die berichteten statistische Verfahren sind nicht zuzuordnen. - Produktivitätskorrelationen wurden mit Korrelation nach Spearman berechnet - Verlorene Arbeitstage während der letzten 6 Wochen wurden mit einem t-Test verglichen - Verlorene Arbeitstage während des Studienverlaufs (statistische Methoden nicht berichtet) Der Umgang mit fehlenden Daten wurde nicht beschrieben.

Anhang 2

Ausgeschlossene Studien mit zum Ausschluss führender Begründung

ausgeschlossene Studien	Auschlussgrund
Affaitati et al. 2009	Die Anforderungen an ein chronisches Geschehen wurden nicht erfüllt.
Al-Smadi et al. 2003	keine arbeitsbezogenen Daten veröffentlicht
Arnold et al. 2011	keine relevanten arbeitsbezogenen Daten veröffentlicht
Bagust et al. 2009	keine arbeitsbezogenen Daten veröffentlicht
Bennell et al. 2007	keine arbeitsbezogenen Daten erhoben
Beurskens et al. 1997	Die Anforderungen an ein chronisches Geschehen wurden nicht erfüllt.
Bjordal et al. 2008	Die Anforderungen an ein chronisches Geschehen wurden nicht erfüllt.
Bombardier und Raboud 1991	keine arbeitsbezogenen Daten veröffentlicht
Borg et al. 1991	keine placebokontrollierte Untersuchung
Braun et al. 2002	keine arbeitsbezogenen Daten veröffentlicht
Bresnihan 2002	Die untersuchten Studien berichteten entweder nicht über Arbeitsendpunkte oder verglichen nicht mit Placebo.
Brox et al. 2008	keine placebokontrollierte Untersuchung
Bruyere et al. 2008	keine arbeitsbezogenen Daten veröffentlicht
Bunzli et al. 2011	Die untersuchten Studien kontrollierten nicht mit Placebo oder die Anforderungen an ein chronisches Geschehen wurden nicht erfüllt.
Busanich und Verscheure 2006	keine arbeitsbezogenen Daten veröffentlicht
Cade et al. 1976	keine arbeitsbezogenen Daten aus der placebokontrollierten Phase berichtet
Cardiel et al. 2010	keine arbeitsbezogenen Daten erhoben
Chen et al. 2006	keine arbeitsbezogenen Daten veröffentlicht
Chen et al. 2013	keine arbeitsbezogenen Daten veröffentlicht
Chiu et al. 2011	keine arbeitsbezogenen Endpunkte untersucht
Chou et al. 2007	Die Anforderungen an ein chronisches Geschehen wurden nicht erfüllt.
Clare et al. 2004	Die untersuchten Studien waren nicht mit Placebo kontrolliert.
Clark et al. 2004	keine arbeitsbezogenen Daten veröffentlicht
Coghlan et al. 2008	Die untersuchten Studien berichteten entweder nicht über Arbeitsendpunkte oder waren nicht mit Placebo kontrolliert.
Cole et al. 2008	keine relevanten arbeitsbezogenen Daten veröffentlicht
Davies et al. 1999	keine arbeitsbezogenen Daten veröffentlicht
Falco et al. 2012	keine arbeitsbezogenen Daten veröffentlicht
Fraser et al. 2005	keine arbeitsbezogenen Daten veröffentlicht
Freeman 2006	Die untersuchten Studien waren entweder keine randomisierten, klinischen, placebokontrollierten Studien oder berichteten nicht über Arbeitsendpunkte.
Furlan et al. 2008	Die untersuchten Studien wurden nicht mit Placebo kontrolliert oder die Anforderungen an ein chronisches Geschehen wurden nicht erfüllt.
Furlan et al. 2010	Die arbeitsbezogenen Studien waren nicht verblindet.
Geisser et al. 2005	keine arbeitsbezogenen Daten veröffentlicht
Gibson und Waddell 2005	keine arbeitsbezogenen Daten veröffentlicht
Goldenberg et al. 1996	keine arbeitsbezogenen Daten veröffentlicht
Han et al. 2008	Es wurde kein Vergleich zwischen den aktiv behandelten und den Kontrollgruppen vorgenommen.
Hennigan et al. 2008	keine vollständig mit Placebo kontrollierte Untersuchung
Hewitson et al. 2000	keine arbeitsbezogenen Daten veröffentlicht
Heymans et al. 2004	keine placebokontrollierte Untersuchungen
Hill und Hill 1975	keine arbeitsbezogenen Daten erhoben
Hudson et al. 2009	keine relevanten arbeitsbezogenen Daten veröffentlicht
Hunter et al. 2009	keine arbeitsbezogenen Daten veröffentlicht
Hutchinson et al. 2012	die arbeitsbezogene Studie kontrollierte nicht mit Placebo
Itoh und Kitakoji 2007	keine arbeitsbezogenen Daten veröffentlicht
Iversen et al. 2011	keine arbeitsbezogenen Daten veröffentlicht

Kay und Rahman 2010	keine arbeitsbezogenen Daten veröffentlicht
Khadilkar et al. 2008	Die relevante Studie dieses Reviews wurde bereits erfasst: Jarzem et al. 2005. Es wurden keine weiteren arbeitsbezogenen Daten berichtet.
Kimel et al. 2011	keine arbeitsbezogenen Daten veröffentlicht
Kirveskari und Alanen 1984	Die Anforderungen an ein chronisches Geschehen wurden nicht erfüllt.
Kuijpers et al. 2011	keine arbeitsbezogenen Daten veröffentlicht
Lerner et al. 2012	keine klinische Untersuchung von Arbeitsendpunkten
Listing et al. 2004	keine placebokontrollierte Untersuchung
Malottki et al. 2011	keine arbeitsbezogenen Daten veröffentlicht
Manheimer et al. 2005	Die untersuchten Studien berichteten entweder nicht über Arbeitsendpunkte oder waren nicht placebokontrolliert.
McLeod et al. 2007	keine arbeitsbezogenen Daten veröffentlicht
Mittendorf et al. 2008	keine placebokontrollierte Untersuchung
Moreland 2004	Die untersuchten Studien berichteten entweder nicht über Arbeitsendpunkte oder waren nicht mit Placebo kontrolliert.
Navarro-Millán et al. 2012	keine arbeitsbezogenen Daten veröffentlicht
Niemistö et al. 2005	keine placebokontrollierte Untersuchung
Nogid und Pham 2006	keine arbeitsbezogenen Daten veröffentlicht
O'Connell et al. 2011	keine arbeitsbezogenen Daten veröffentlicht
Osterhaus et al. 2009	Es wurde kein Vergleich zwischen den aktiv behandelten Gruppen und den Kontrollgruppen vorgenommen.
Pach et al. 2011	keine arbeitsbezogenen Daten veröffentlicht
Pope et al. 2004	keine arbeitsbezogenen Daten veröffentlicht
Price et al. 2005	Daten von Patienten mit akuten und chronsichen Schmerzen wurden vermischt.
Rodgers et al. 2011	keine arbeitsbezogenen Daten veröffentlicht
Rogvi-Hansen et al. 1991	Die Anforderungen an ein chronisches Geschehen wurden nicht erfüllt.
Rossini et al. 2007	keine arbeitsbezogenen Daten veröffentlicht
Rubinstein et al. 2011	Die untersuchten Studien berichteten nicht über arbeitsbezogene Endpunkte oder erfüllten nicht die Anforderungen an ein chronsiches Geschehen.
Ruiz Garcia et al. 2011	keine arbeitsbezogenen Daten veröffentlicht
Schmitz et al. 2012	keine arbeitsbezogenen Daten veröffentlicht
Schneider et al. 2012	keine arbeitsbezogenen Daten veröffentlicht
Scott 1999	keine arbeitsbezogenen Daten veröffentlicht
Skljarevski et al. 2010c	Open Label Erweiterung der Studie von Skljarevski et al. 2010a
Skljarevski et al. 2011	keine arbeitsbezogenen Daten veröffentlicht
Skljarevski et al. 2012	keine arbeitsbezogenen Daten veröffentlicht
Sörensen et al. 1996	keine placebokontrollierte Untersuchung
Stacey und Glanzman 2003	keine arbeitsbezogenen Daten veröffentlicht
Szczurko et al. 2009	Die Anforderungen an ein chronisches Geschehen wurden nicht erfüllt.
Thompson et al. 1988	keine arbeitsbezogenen Daten veröffentlicht
Turner et al. 1995	kein randomisierten, klinischen Studien erfasst
Wernicke et al. 2007	keine placebokontrollierte Untersuchung
Winemiller et al. 2005	Die Anforderungen an ein chronisches Geschehen wurden nicht erfüllt.
Woolacott et al. 2006	keine arbeitsbezogenen Daten veröffentlicht
Yang et al. 2012	keine arbeitsbezogenen Daten veröffentlicht
Zanette et al. 2008	keine arbeitsbezogenen Daten veröffentlicht

Anhang 3
Erfasste Erkrankungen

Englischer Suchbegriff	Erfasste Erkrankungen	deutsche Übersetzung	Kurzbeschreibung	chronische Schmerzen	
osteoarthritis	osteoarthritis Synonyme: osteoarthrosis degenerative joint disease	Arthrose/Osteoarthrose	Beschreibt eine Gruppe von Erkrankungen, die Gelenksymptome, Beschädigungen am Gelenkknorpel und Veränderungen an benachbarten Knochenstrukturen aufweisen.	Gelenkschmerzen gehen häufig mit Osteoarthrose einher.	1
arthritis		Arthritis	Entzündliche Erkrankung eines Gelenkes		
	rheumatoid arthritis	rheumatoide Arthritis	RA beschreibt eine systemische, häufig chronisch und voranschreitende Bindegewebserkrankung. Die Gelenke sind dabei immer betroffen. Sehnen, Sehnenscheiden, Schleimbeutel, Augen und anderer Organbefall ist möglich.	Neben Steifigkeit, Schwellungen und Bewegungseinschränkungen sind Schmerzen fester Teil der Symptomatik.	3
	psoriatic arthritis	Psoriasisarthritis	PA erfasst eine Gruppe von Patienten, welche Arthritis und Psoriasis aufweisen. Weitere Diagnosekriterien können sein: Daktylitis, Nagelauffälligkeiten, familiäres Auftreten von Psoriasis.	Gelenkschmerzen sind zentraler Teil der Symptomatik	3
	reactive arthritis	reaktive Arthritis	Arthritis, die nach einer Infektion des Gastrointestinal- oder Urogenitaltraktes auftritt.	Fersenschmerzen oder tiefe Rückenschmerzen sind häufige Begleiterscheinungen.	3
Reiter's syndrome	Reiter's syndrome	Morbus Reiter	Besondere Form der reaktiven Arthritis. Symptomtrias: Arthritis, Urethritis und Konjunktivitis	siehe reaktive Arthritis	3
spondylarthritis		Spondarthritiden	Sammelbegriff für entzündliche, rheumatiode Erkrankungen, die die Wirbelsäule und die Iliosakralgelenke befallen.		
spondylitis		Spondylitis	Entzündung der Wirbelkörper		
	ankylosing spondylitis	Spondylitis ankylosans	Chronische entzündliche Erkrankung, die sich neben destruktiven auch durch proliferierende Vorgänge in den befallenen Gelenken auszeichnet. Dies kann bis zur vollständigen Versteifung der Gelenke führen. Befallen sind oft Iliosakralgelenke und die Wirbelsäule.	Zu den typischen Symptomen zählen tiefe Rückenschmerzen, die bis in die Wade ziehen können.	3

granulomatosis polyangiitis	Wegener's granulomatosis Synonym: granulomatosis with polyangiitis	Wegener-Granulomatose	Vaskulitis, die mit nekrotisierenden und granulomatösen Veränderungen an Arterien und Venen einher geht.	Besonders bei jungen Patienten werden Arthralgien und Synovitiden beobachtet.	3
panarteritis nodosa polyarteritis periarteritis	panarteritis nodosa polyarteritis nodosa periarteritis nodosa	Panarteritis nodosa Polyarteritis nodosa Periarteritis nodosa	Generalisierte, nekrotisierende Gefäßentzündung, die die gesamte Gefäßwand befällt. Je nach Lokalisation und betroffenen Organen unterscheidet sich die Symptomatik.	Arthralgien und Synovitiden sind möglich. Bei ca. 2/3 der Patienten kommen Myalgien vor. Mono- oder Polyneuritis können ebenfalls vorkommen.	3
granulomatosis Churg–Strauss syndrome	allergic granulomatosis Synonym: Churg–Strauss syndrome	Churg–Strauss Syndrom	Hat eine ähnliche Symptomatik wie die Panarteritis nodosa. Eine Lungenbeteiligung, eine Eosinophilie und granulomatöse Infiltrate in und um die Gefäßewände herum erlauben eine Differenzierung.	Siehe Panarteritis nodosa	3
polyangiitis polyarteritis	microscopic polyangiitis microscopic polyarteritis	Mikroskopische Polyangiitis	Sonderform der Panarteritis nodosa, die kleine Blutgefäße befällt.	Länger anhaltende Arthralgien und Myalgien sind häufig.	3
polymyalgia rheumatica	polymyalgia rheumatica	Polymyalgia rheumatica	Entzündliche Systemerkrankung, die besonders ältere Menschen betrifft. Histologisch lässt sich oft eine Riesenzellarteriitis feststellen.	Typische Beschwerden sind proximale, symmetrische Muskelschmerzen	3
connective tissue disease	connective tissue disease	Kollagenose	Entzündliche Systemerkrankungen des Bindegewebes		
systemic lupus erythematosus	systemic lupus erythematosus	Systemischer Lupus erythematodes	Klassische Autoimmunerkrankung,die verschiedene Orange befällt und oft mit Synovitiden einher geht.	Gelenkschmerzen finden sich bei 90% der Erkrankten.	3
polymyositis	polymyositis	Polymyositis	Gruppe systemischer Bindegewebserkrankung mit ungeklärter Ätiologie, die besonders durch eine Entzündung der Skelettmuskulatur in Erscheinung tritt.	Muskelkater ähnliche Myalgien sind häufig, während Arthralgien seltener vorkommen.	3
myofascial pain syndrome	myofascial pain syndrome	Myofasziales Schmerzsyndrom	Erfasst alle Krankheitskomplexe, die Schmerzen des Bewegungsapparates aufweisen, aber nicht innerhalb der Gelenkkapsel oder des Periosts beheimatet sind. Die Ursache der Schmerzen ist nicht auf eine Rheumatisch-entzündliche oder neurologische Systemerkrankung zurück zu führen.	Ein Großteil der Patienten mit chronischen Schmerzen haben ein Myofasziales Schmerzsyndrom.	3

fibromyalgia	fibromyalgia	Fibromyalgie	Generalisiertes Schmerzsyndrom mit ungeklärter Ursache. Zur wissenschaftlichen Erfassung einer Patientenpopulation sollten mindestens 11 von 18 definierten Tender Points auf Palpation mit Schmerz reagieren. Diese sind über den Körper verteilt (meist im Bereich der Sehnenansätze oder Muskel-Sehnen-Übergängen) und sollen so die Schmerzgeneralisierung widerspiegeln. Klinisch sind noch weitere vegetative und funktionelle Störungen relevant.	Die Erkrankung zeichnet sich durch generalisierte Schmerzsymptomatik aus.	3
facet syndrome	facet syndrome	Facettensyndrom	Bezeichnet Schmerzen der Wirbelsäule, die hauptsächlich von den Wirbelbogengelenken verursacht werden.	Bei dem Syndrom werden eng beschriebene oberflächliche Schmerzen angegeben.	3
gout	gout	Gicht	Umfasst Veränderungen des Bewegungsapparates, die Folge einer Störung des Proteinstoffwechsels sind. Ausgelöst durch exogene Noxen steigt die Harnsäurekonzentration in den Körperflüssigkeiten bis zur Ausfällung von Uratkristallen an. Folgen sind unter anderem Synovitiden/Arthritis urica.	In der akuten und chronischen Phase der Arthritis urica wird über heftige Gelenkbeschwerden berichtet.	3
chronic whiplash injury	chronic whiplash injury	chronisches Schleudertrauma	Energieübertragung auf die Halswirbelsäule bei Beschleunigungs- oder Abbremsvorgängen kann zu klinischen Symptomen führen.	Primäre Symptome sind Nackenschmerzen und Kopfschmerzen.	4
complex regional pain syndrome	complex regional pain syndrome	Komplexes regionales Schmerzsyndrom	Beschreibt eine Vielzahl von schmerzhaften Erkrankungen, die nach einer Verletzung auftreten und durch unerwartete Stärke oder Dauer der Symptomatik auffallen. Oft sind Einschränkungen der Motorik und ein progressiver Verlauf zu beobachten.	Schmerzen zählen zu den primären Symptomen dieser Erkrankungen.	5
chronic pelvic pain syndrome	chronic pelvic pain syndrome	chronische Beckenschmerzen	Unscharf umschriebenes chronisches Schmerzsyndrom im Beckenbereich. Häufig tritt es gleichzeitig mit klarer definierten Befunden auf, die jedoch die erfolglose Behandlung der Schmerzen nicht rechtfertigen. Es sind sowohl Männer als auch Frauen von der Erkrankung betroffen.	Chronische Schmerzen sind eines der Hauptsymptome.	2

1 Altman et al. 1986
2 Baranowski et al. 2014
3 Hettenkofer et al. 2003
4 Spitzer et al. 1995
5 Stanton-Hicks et al.1995

Anhang 4

Sucheingaben und ihre korrespondierenden Aufarbeitungen vom „Automatic-Term-Mapping“

work	"work"[MeSH Terms] OR "work"[All Fields]
sick leave	"sick leave"[MeSH Terms] OR ("sick"[All Fields] AND "leave"[All Fields]) OR "sick leave"[All Fields]
economic	"economics"[MeSH Terms] OR "economics"[All Fields] OR "economic"[All Fields]
employment	"employment"[MeSH Terms] OR "employment"[All Fields]
labor	"labour"[All Fields] OR "work"[MeSH Terms] OR "work"[All Fields] OR "labor"[All Fields] OR "labor, obstetric"[MeSH Terms] OR ("labor"[All Fields] AND "obstetric"[All Fields]) OR "obstetric labor"[All Fields]
osteoarthritis	"osteoarthritis"[MeSH Terms] OR "osteoarthritis"[All Fields]
osteoarthrosis	"osteoarthritis"[MeSH Terms] OR "osteoarthritis"[All Fields] OR "osteoarthrosis"[All Fields] OR "joint diseases"[MeSH Terms] OR ("joint"[All Fields] AND "diseases"[All Fields]) OR "joint diseases"[All Fields]
spondylarthritis	"spondylarthritis"[MeSH Terms] OR "spondylarthritis"[All Fields]
arthritis	"arthritis"[MeSH Terms] OR "arthritis"[All Fields]
Reiter's syndrome	"arthritis, reactive"[MeSH Terms] OR ("arthritis"[All Fields] AND "reactive"[All Fields]) OR "reactive arthritis"[All Fields] OR ("reiter's"[All Fields] AND "syndrome"[All Fields]) OR "reiter's syndrome"[All Fields]
gout	"gout"[MeSH Terms] OR "gout"[All Fields]
polymyalgia rheumatica	"polymyalgia rheumatica"[MeSH Terms] OR ("polymyalgia"[All Fields] AND "rheumatica"[All Fields]) OR "polymyalgia rheumatica"[All Fields]
spondylitis	"spondylitis"[MeSH Terms] OR "spondylitis"[All Fields]
polyangiitis	"systemic vasculitis"[MeSH Terms] OR ("systemic"[All Fields] AND "vasculitis"[All Fields]) OR "systemic vasculitis"[All Fields] OR "polyangiitis"[All Fields]
joint	"joints"[MeSH Terms] OR "joints"[All Fields] OR "joint"[All Fields]
pain	"pain"[MeSH Terms] OR "pain"[All Fields]
polymyositis	"polymyositis"[MeSH Terms] OR "polymyositis"[All Fields]
connective tissue disease	"connective tissue diseases"[MeSH Terms] OR ("connective"[All Fields] AND "tissue"[All Fields] AND "diseases"[All Fields]) OR "connective tissue diseases"[All Fields] OR ("connective"[All Fields] AND "tissue"[All Fields] AND "disease"[All Fields]) OR "connective tissue disease"[All Fields]
systemic lupus erythematosus	"lupus erythematosus, systemic"[MeSH Terms] OR ("lupus"[All Fields] AND "erythematosus"[All Fields] AND "systemic"[All Fields]) OR "systemic lupus erythematosus"[All Fields] OR ("systemic"[All Fields] AND "lupus"[All Fields] AND "erythematosus"[All Fields])
panarteritis nodosa	"polyarteritis nodosa"[MeSH Terms] OR ("polyarteritis"[All Fields] AND "nodosa"[All Fields]) OR "polyarteritis nodosa"[All Fields] OR ("panarteritis"[All Fields] AND "nodosa"[All Fields]) OR "panarteritis nodosa"[All Fields]
periarteritis nodosa	"polyarteritis nodosa"[MeSH Terms] OR ("polyarteritis"[All Fields] AND "nodosa"[All Fields]) OR "polyarteritis nodosa"[All Fields] OR ("periarteritis"[All Fields] AND "nodosa"[All Fields]) OR "periarteritis nodosa"[All Fields]
Churg-Strauss syndrome	"churg-strauss syndrome"[MeSH Terms] OR ("churg-strauss"[All Fields] AND "syndrome"[All Fields]) OR "churg-strauss syndrome"[All Fields] OR ("churg"[All Fields] AND "strauss"[All Fields] AND "syndrome"[All Fields]) OR "churg strauss syndrome"[All Fields]

myofascial pain syndrome	"myofascial pain syndromes"[MeSH Terms] OR ("myofascial"[All Fields] AND "pain"[All Fields] AND "syndromes"[All Fields]) OR "myofascial pain syndromes"[All Fields] OR ("myofascial"[All Fields] AND "pain"[All Fields] AND "syndrome"[All Fields]) OR "myofascial pain syndrome"[All Fields] OR "fibromyalgia"[MeSH Terms] OR "fibromyalgia"[All Fields] OR ("myofascial"[All Fields] AND "pain"[All Fields] AND "syndrome"[All Fields])
pelvic pain	"pelvic pain"[MeSH Terms] OR ("pelvic"[All Fields] AND "pain"[All Fields]) OR "pelvic pain"[All Fields]
syndrome	"syndrome"[MeSH Terms] OR "syndrome"[All Fields]
back pain	"back pain"[MeSH Terms] OR ("back"[All Fields] AND "pain"[All Fields]) OR "back pain"[All Fields]
neck pain	"neck pain"[MeSH Terms] OR ("neck"[All Fields] AND "pain"[All Fields]) OR "neck pain"[All Fields]
fibromyalgia	"fibromyalgia"[MeSH Terms] OR "fibromyalgia"[All Fields]
whiplash injury	"whiplash injuries"[MeSH Terms] OR ("whiplash"[All Fields] AND "injuries"[All Fields]) OR "whiplash injuries"[All Fields] OR ("whiplash"[All Fields] AND "injury"[All Fields]) OR "whiplash injury"[All Fields]
neuralgia	"neuralgia"[MeSH Terms] OR "neuralgia"[All Fields]
complex regional pain syndrome	"complex regional pain syndromes"[MeSH Terms] OR ("complex"[All Fields] AND "regional"[All Fields] AND "pain"[All Fields] AND "syndromes"[All Fields]) OR "complex regional pain syndromes"[All Fields] OR ("complex"[All Fields] AND "regional"[All Fields] AND "pain"[All Fields] AND "syndrome"[All Fields]) OR "complex regional pain syndrome"[All Fields]
placebo	"placebos"[MeSH Terms] OR "placebos"[All Fields] OR "placebo"[All Fields]
sham	"salicylhydroxamic acid"[Supplementary Concept] OR "salicylhydroxamic acid"[All Fields] OR "sham"[All Fields]

Publikationen im Zusammenhang mit dieser Dissertation

Wolf I, Friede T, Hallier E, Straube S (2012): Beschreibung arbeitsbezogener Endpunkte in Studien zur Behandlung chronischer Schmerzerkrankungen – systematische Literaturübersicht. in: 52.Wissenschaftliche Jahrestagung der Deutschen Gesellschaft für Arbeitsmedizin und Umweltmedizin 14. - 17. März 2012, Göttingen. Dokumentation, 460-463

Wolf I, Friede T, Hallier E, Straube S (2014): Arbeitsbezogene Studienendpunkte in randomisierten placebokontrollierten Therapiestudien zu chronischen Schmerzerkrankungen – Systematischer Review und Meta-Analyse. Umweltmedizin – Hygiene – Arbeitsmedizin 19, 211-212

Wolf I, Friede T, Hallier E, Straube S (2014): Work-related outcomes in randomised placebo-controlled pain trials: a systematic review and meta-analysis. J Occup Med Toxicol 9:25

Danksagung

Für die engagierte Betreuung vor und während der Dissertationserstellung spreche ich Herrn Prof. Dr. Sebastian Straube sowie Herrn Prof. Dr. Tim Friede meinen herzlichen Dank aus. Ihre schnellen Reaktionszeiten, auch außerhalb üblicher Arbeitszeiten, ermöglichten ein motiviertes und produktives Bearbeiten des komplexen Themas. Die Teilhabe an ihrer Erfahrung und ihrer gewissenhaften Arbeitsmethodik waren für mich von unschätzbarem Nutzen.