

AUS DER KLINIK FÜR PSYCHIATRIE UND
PSYCHOTHERAPIE DER MEDIZINISCHEN FAKULTÄT
CHARITÉ – UNIVERSITÄTSMEDIZIN BERLIN

DISSERTATION

**Wirksamkeit stationärer Psychotherapie bei
Depressionen - eine Metaanalyse
kontrollierter Studien**
*Efficacy of inpatient psychotherapy for major
depressive disorder: a meta-analysis of
controlled trials*

*zur Erlangung des akademischen Grades
Doctor medicinae (Dr. med.)*

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von
Cora SCHEFFT
aus Göttingen

Datum der Promotion: 21.Juni 2020

Inhaltsverzeichnis

Abkürzungsverzeichnis	v
Zusammenfassung	vii
Abstract	ix
1 Einführung	1
1.1 Unipolare Depression	1
1.2 Therapie der unipolaren Depression	1
1.2.1 Inanspruchnahme und Kosten der stationären Depressionsbehandlung	2
1.2.2 Psychotherapie der unipolaren Depression	2
1.2.3 Psychotherapie in randomisiert kontrollierten Studien	3
1.2.4 Wirksamkeit von Psychotherapie im ambulanten und stationären Rahmen	6
1.3 Fragestellung und Hypothesen	7
2 Methodik	9
2.1 Auswahl der Methode	9
2.2 Festlegung der Ein- und Ausschlusskriterien	9
2.3 Datenbankrecherche und Studienselektion	10
2.4 Datenextrahierung	10
2.5 Metaanalyse	11
2.5.1 Standardisierte Mittelwertsdifferenz	11
2.5.2 Random effects model	12
2.5.3 Heterogenität	13
2.5.4 Subgruppen- und Sensitivitätsanalysen	14
2.5.5 Ermittlung des Publikationsbias	14
2.6 Qualitätsbeurteilung	15
3 Ergebnisse	17
3.1 Ergebnisse der Hauptmetaanalyse	17
3.1.1 Ergebnisse der Subgruppenanalyse	17
3.1.2 Ergebnisse der Tests auf Publikationsbias	18

3.2	Ergebnisse der Qualitätsbeurteilung	19
4	Diskussion	21
4.1	Bedeutung der Ergebnisse	21
4.2	Qualitätsaspekte und Methodik	23
4.3	Limitationen und Ausblick	24
	Literatur	25
	Eidesstattliche Erklärung	29
	Ausführliche Anteilserklärung an der erfolgten Publikation	31
	Auszug aus der Journal Summary List	33
A	Originalpublikation und Supporting Information	35
A.1	Schefft C, Guhn A, Brakemeier EL, Sterzer P, und Köhler S. „Efficacy of inpatient psychotherapy for major depressive disorder: a meta-analysis of controlled trials“. In: <i>Acta Psychiatrica Scandinavica</i> 139.4 (2018), S. 322–335. https://doi.org/10.1111/acps.12995	35
A.2	Supporting Information	51
A.2.1	Table S1: PRISMA 2009 Checklist	51
A.2.2	Appendix S1: Search protocol	56
A.2.3	Appendix S2: Supplemental Figures S1–S4	59
B	Lebenslauf	65
C	Publikationsliste	69
	Danksagung	71

Abkürzungsverzeichnis

BDI	Beck Depression Inventory
CBASP	Cognitive Behavioral Analysis System of Psychotherapy
DSM	Diagnostic and Statistical Manual of Mental Disorders
FEM	Fixed effect model
H	Hypothese
ICD-10	International Statistical Classification of Diseases and Related Health Problems
IPT	Interpersonelle Therapie
ITT	Intention - to - treat
Kap.	Kapitel
KVT	Kognitive Verhaltenstherapie
NNT	Number needed to treat
MDD	Major depressive disorder
PDT	Psychodynamische Therapien
QIDS-SR	Quick Inventory of Depressive Symptomatology - self-report
RCT	Randomized controlled trial
REM	Random effects model
S.	Seite
SD	Standardabweichung
SMD	Standardisierte Mittelwertsdifferenz

Zusammenfassung

Fragestellung: In dieser Metaanalyse soll die Wirksamkeit stationärer Psychotherapie depressiver Störungen im Vergleich zu Kontrollbedingungen ohne psychotherapeutische Intervention beurteilt werden. Dabei werden Faktoren wie Art der Kontrollbedingung, die Verwendung von Fremd- bzw. Selbstbeurteilungsskalen und die Anzahl an Therapieeinheiten als moderierende Variablen berücksichtigt.

Methode: Eine systematische Suche in vier biomedizinischen Datenbanken (CENTRAL, PsycINFO, MEDLINE, Web of Science) wurde unternommen. Als primärer Endpunkt galt die Differenz der Depressionsschwere zwischen der Interventions- und Kontrollbedingung ermittelt durch Selbstbeurteilungsskalen, die in einer *random effects* Metaanalyse zusammengefasst wurden. In Subgruppenanalysen wurden Studien mit verschiedenen Arten von Kontrollbedingungen verglichen, wobei unterschieden wurde zwischen a) keine Behandlung (z.B. Warteliste), b) *treatment as usual* (TAU) - nicht durch die Studie standardisiert und c) TAU - manualisiert, durch die Studie standardisiert. Ferner wurden Studien mit ≤ 10 , $11 - 20$ und ≥ 21 Therapieeinheiten, sowie Ergebnisse von Selbst- und Fremdbeurteilungsmaßen verglichen. Eine Prüfung auf einen Publikationsbias der Metaanalyse und eine Bewertung des Risikos systematischer Verzerrungen innerhalb der einzelnen Studien wurde vorgenommen.

Resultate: Es wurden 14 relevante Studien ($k = 19$ Vergleiche) mit insgesamt 1080 Patienten gefunden. Die Metaanalyse wies eine kleine, statistisch signifikante Überlegenheit der stationären Psychotherapie gegenüber Kontrollbedingungen ohne Psychotherapie aus ($g = 0.24$, $p < 0.001$, $I^2 = 0\%$, $NNT = 7.4$). Die Effekte der Intervention zeigten sich über eine Periode von mindestens 12 Monaten stabil ($k = 9$, $n = 710$, $g = 0.21$, $p = 0.04$, $I^2 = 30\%$). Im Subgruppenvergleich von stationärer Psychotherapie mit standardisierten Kontrollbedingungen zeigte sich keine Überlegenheit der Psychotherapie. Im Vergleich zu Warteliste oder unstandardisiertem TAU blieb der Effekt bestehen. Eine höhere Anzahl an Therapiesitzungen zeigte tendenziell geringere Effekte. Ergebnisse von Fremd- und Selbstbeurteilungsmaßen unterschieden sich marginal. Es fand sich keine Evidenz für einen Publikationsbias.

Schlussfolgerung: Ein kleiner jedoch robuster und anhaltender Effekt stationärer Psychotherapie gegenüber einer gewöhnlichen stationären Krankenhausbehandlung konnte für depressive Störungen gezeigt werden. Der Effekt war in einer Subgruppe von Studien mit stärker standardisierten Kontrollbedingungen nicht mehr nachzuweisen.

Abstract

Objective: This meta-analysis aims to investigate the efficacy of inpatient psychotherapy of depressive disorders compared to control conditions without psychotherapy. As moderating factors, we consider the type of control condition, use of self- vs clinician ratings and number of treatment sessions.

Methods: A systematic search of four biomedical databases (CENTRAL, PsycINFO, MEDLINE, Web of Science) was performed. The differences in self-rated depression severity between intervention and control groups served as primary outcomes, which were synthesized in a random effects meta-analysis. In subgroup analyses, studies were compared based on differences in the type of control conditions they applied. We distinguished a) no treatment or waitlist controls, b) treatment as usual (TAU) - not standardized by trial design, c) TAU - standardized by trial design, manualized. Subgroups of trials providing ≤ 10 , $11 - 20$ or ≥ 21 treatment sessions were compared as well as results from self- versus clinician ratings.

Results: Fourteen relevant studies ($k = 19$ comparisons) with a total of 1080 patients were retrieved. The meta-analysis showed a small, statistically significant superiority of inpatient psychotherapy compared to control conditions without psychotherapy ($g = 0.24$, $p < 0.001$, $I^2 = 0\%$, $NNT = 7.4$). The effects of the intervention were stable over a minimum 12-months follow-up period ($g = 0.21$, $p = 0.04$, $I^2 = 30\%$). In subgroup comparisons of inpatient psychotherapy to manualized and standardized control conditions, no superiority of psychotherapy was shown. Compared to waitlist and non-standardized TAU the significant effect remained. A higher number of treatment sessions showed trendwise smaller effects. Results between self- and clinician ratings differed only marginally. There was no evidence of a publication bias.

Conclusions: We found evidence for a small but robust and lastig effect of inpatient psychotherapy in patients with depressive disorders when compared to treatment as usual. The effect was not present in a subgroup of studies using more standardized control conditions.

Kapitel 1

Einführung

1.1 Unipolare Depression

Die unipolare Depression bildet die häufigste Verlaufsform unter den affektiven Störungen. Sie ist gekennzeichnet durch einen meist episodischen Verlauf mit den Kernsymptomen niedergeschlagener Stimmung, Verlust von Interesse und Motivation oder Antriebsverlust. Dazu können erhöhte Ermüdbarkeit, Schlaf- und Appetitstörungen, vermindertes Selbstwertgefühl, übermäßige Schuldgefühle, Konzentrationsbeschwerden oder Verlust der Libido auftreten. Die geschätzten Prävalenzen unipolar depressiver Störungen variieren zwischen Regionen [1]. In Deutschland beträgt die Lebenszeitprävalenz der unipolaren Depression ca. 12% [2], in den USA bis zu 21% [3]. Unipolare Depressionen beeinträchtigen erheblich die Lebensqualität der Betroffenen, sowohl durch Verschlechterung des individuellen Befindens, als auch durch Einschränkung sozialer und ökonomischer Teilhabe. Basierend auf Daten der Weltgesundheitsorganisation wurden für 2030 die unipolar depressive Störungen als weltweit zweithäufigste Ursache für die Akkumulation von *disability adjusted life years* (DALY's) – Lebensjahre die krankheitsbedingt beeinträchtigt oder verloren werden – vorhergesagt [4].

1.2 Therapie der unipolaren Depression

Die Therapie der unipolaren Depression gliedert sich in die Phasen der Akut- und Erhaltungstherapie sowie der Rezidivprophylaxe. Das Ziel der Akuttherapie sollte eine Vollremission der Symptomatik sein [5]. Psychotherapie kann in der Akuttherapie eingesetzt werden und wird zur Erhaltungstherapie und Rezidivprophylaxe nach initialer Remission empfohlen [5]. Unterschiede in den Behandlungsempfehlungen orientieren sich am Schweregrad der Depression: Leichtgradigen depressiven Episoden oder depressiven Verstimmungen kann zunächst mit einer aktiv abwartenden Haltung oder niederschweligen supportiven Interventionen begegnet werden. Ferner können leicht- und mittelgradige depressive Episoden entweder pharmakologisch oder psychotherapeutisch behandelt werden; schwere Depressionen sollten hingegen mit einer Kombination aus beidem behandelt werden [5]. Der

Pharmakotherapie kommt aufgrund ihrer breiteren Verfügbarkeit gegenüber der Psychotherapie eine hervorgehobene Stellung zu. Insgesamt zeigt die Pharmakotherapie der Depression eine mäßige Wirksamkeit und eine nahezu gleich hohe Abbruchwahrscheinlichkeit gegenüber der Placebobehandlung [6].

Die Versorgung von Patienten, die an einer depressiven Störung leiden, entfällt zu einem großen Teil auf Allgemeinarztpraxen. Ferner auf die entsprechenden Fachärzte, psychologische und ärztliche Psychotherapeuten und die stationäre psychiatrisch- bzw. psychosomatisch - psychotherapeutische Behandlung. Indikationen für eine stationäre Behandlung umfassen akute Suizidalität und fehlende Absprachefähigkeit, Belastungsfaktoren in der Häuslichkeit, fehlgeschlagene Versuche einer ambulanten Behandlung, psychotische Symptome oder Folgeerscheinungen von Suizidversuchen [5].

1.2.1 Inanspruchnahme und Kosten der stationären Depressionsbehandlung

Einer auf Krankenkassendaten basierenden Schätzung zufolge entfallen 44% der Behandlungskosten unipolar depressiver Störungen auf die stationäre Behandlung, die jedoch nur von 6% der Betroffenen in Anspruch genommen wird [7]. Bei Patienten mit schweren Depressionen entfallen ca. 75% der jährlichen Behandlungskosten auf die stationäre Therapie [7]. Zu ähnlichen Zahlen kommt eine retrospektive Analyse von 18139 Versicherten, die im Jahre 2010 aufgrund einer depressiven Störung erstbehandelt wurden. Hierbei wurden 4.6% der Patienten stationär behandelt, wodurch 51.1% der Gesamtkosten verursacht wurden, was einer Summe von EUR 8699 (SD 7321) pro stationärem Patient entsprach [8]. Die Wahrscheinlichkeit, aufgrund einer Depression stationär behandelt zu werden, war laut dieser Studie vor allem erhöht, wenn psychiatrische Komorbiditäten oder eine schwere Symptomatik vorlagen [8]. Eine Erhebung an 322 Patienten, die im Jahre 2016 stationär in der Universitätsklinik Kiel mit den Diagnosen unipolar depressiver Störungen behandelt wurden, schätzte ähnlich der Analyse aus 2010 die durchschnittlichen Kosten der Behandlung pro Patient auf EUR 8845 (SD 428). Die durchschnittliche Liegedauer betrug in dieser Studie ca. 37 Tage (SD 1.8) [9]. Eine für Deutschland repräsentative Erhebung ergab, dass die überwiegende Zahl von stationär behandelten Patienten mit depressiven Störungen eine Kombination aus Psychotherapie und Pharmakotherapie erhalten [10]. In dieser Erhebung wurde Psychotherapie jedoch als mindestens eine Sitzung Gruppen- oder Psychotherapie operationalisiert.

1.2.2 Psychotherapie der unipolaren Depression

Eine Kurzdefinition von Psychotherapie könnte lauten: Einem Psychotherapieverfahren liegen spezifische theoretische Annahmen über ein Störungsbild zugrunde, aus denen Interventionstechniken zur Erreichung vorgesehener Behandlungsergebnisse abgeleitet werden [5]. Psychotherapeutische Interventionen haben in der zweiten Hälfte des 20. Jahrhunderts

zunehmende Verbreitung gefunden und somit Interesse am experimentellen Nachweis ihrer Wirksamkeit geweckt [11]. Bei der Therapie unipolar depressiver Störungen finden kognitive Verhaltenstherapie (KVT), psychodynamische Verfahren (PDT, tiefenpsychologisch fundierte oder psychoanalytische Psychotherapie), Interpersonelle Psychotherapie (IPT), Gesprächspsychotherapie oder systemische Psychotherapie Anwendung. Von der gesetzlichen Krankenversicherung werden in Deutschland aktuell ambulante Therapien in den ersten beiden Verfahren erstattet. In den letzten zwei Jahrzehnten wurden die klassischen Psychotherapieverfahren weiterentwickelt und in ihrer Spezifität an die Störungsbilder angepasst. Der überwiegende Teil der Patienten, die Psychotherapie in Anspruch nehmen, tut dies bei einem ambulant niedergelassenen Psychotherapeuten in regelmäßigen Einzel- oder Gruppensitzungen. Hierbei übersteigt jedoch die Nachfrage das Angebot und häufig ist die Aufnahme einer ambulanten Psychotherapie für die Patienten mit mehrmonatigen Wartezeiten verbunden. Im stationären Bereich ist es ebenso möglich und empfohlen [5], Psychotherapie anzubieten. Hierbei kann es sich um ein allgemeines, unspezifisches Angebot handeln, das sich an Patienten verschiedener Diagnosegruppen richtet oder um diagnosespezifische Programme. In beiden Fällen kommen sowohl die oben genannten Verfahren, insbesondere KVT, zum Einsatz, als auch weitere, speziell für Depressionen angepasste Therapien wie z.B. Verhaltensaktivierung, achtsamkeitsbasierte Verfahren, IPT oder CBASP (*Cognitive Behavioral Analysis System of Psychotherapy*). Im stationären Setting wird den Patienten zusätzlich zu Psycho- und Pharmakotherapie zudem oft ein komplementäres Angebot aus Ergo-, Kunst- oder Musiktherapie, Entspannungsverfahren, Psychoedukation, Wach- oder Lichttherapie sowie Elektrokonvulsionstherapie bereitgestellt.

1.2.3 Psychotherapie in randomisiert kontrollierten Studien

Im Rahmen der evidenzbasierten Medizin wird auch an die Psychotherapie der Anspruch erhoben, ihre Wirksamkeit in kontrollierten, randomisierten Studien unter Beweis zu stellen. Die Rationale eines *randomized controlled trials* (RCT) ist, dass kausale Schlüsse über die Wirksamkeit einer Intervention gezogen werden können, indem sie gegen eine Kontrollbedingung *ceteris paribus* verglichen wird. Die Kontrollgruppe erhält in Medikamentenstudien dafür ein Placebo. Idealerweise imitiert das Placebo die Nebenwirkungen des Verums und gleicht ihm in allen Aspekten mit Ausnahme des zu untersuchenden Wirkstoffes. Ein RCT sollte zudem doppelblind durchgeführt werden, sodass weder Patienten, noch diejenigen, die sie behandeln oder die, die die Wirkung der Intervention messen, wissen, ob ein Patient das Placebo oder das Verum erhält. Die Randomisierung dient der Minimierung des Risikos einer systematischen Verzerrung durch Selektionseffekte. Unterschiede in demographischen Variablen zwischen Patienten sollten sich bei erfolgreicher Randomisierung herausmitteln, sodass diese nicht mit den Ergebnissen kovariieren. Interventionen wie die

Psychotherapie unterliegen bei ihrer Prüfung im RCT jedoch bestimmten Besonderheiten, die eine Gefahr für die interne Validität der Studien darstellen [12]:

- Verblindung des Klinikers: Der Psychotherapeut kann nicht für die Therapie, die er durchführt verblindet werden. Dadurch steigt das Risiko für einen *performance bias*. Das bedeutet, dass die Vergleichsgruppen unterschiedlich stark profitieren, weil eingesetzte Therapeuten die Intervention mit größerer Überzeugung und Motivation durchführen als die Kontrollbehandlung. Dadurch würde die Effektivität überschätzt.
- Risiko einer Verzerrung durch *therapeutic allegiance* oder *investigator allegiance*: Dies bezeichnet eine Überschätzung eines Interventionseffekts in einer Studie dadurch, dass die Therapeuten bzw. Studienleiter Experten oder Manualautoren des interessierenden Verfahrens sind und deshalb die Studie nicht ergebnisoffen und neutral gestalten.
- Verblindung der Probanden: Ebenso eingeschränkt ist die Verblindung auf Seiten des Patienten. Der Proband weiß in der Regel, ob er eine Psychotherapie erhält oder nicht, insbesondere wenn die Kontrollbedingung beinhaltet, auf einer Warteliste zu stehen oder eine rein medikamentöse Behandlung zu erhalten. Daraus ergibt sich ebenfalls ein erhöhtes Risiko für eine Überschätzung der Interventionseffekte durch *performance bias*. Ein möglicher Ausweg hierbei ist, eine Kontrollbedingung zu wählen, wobei die Kontrollgruppe entweder eine andere, etablierte Therapie erhält oder eine Behandlung mit unspezifischen Therapiekomponenten. Eine solche Kontrollbehandlung sollte, ähnlich der Placebotablette, in Form und Umfang der Intervention gleichen, ohne dass sie die angenommenen wirksamen Faktoren enthält.

Aufgrund dieser inhärenten Risiken für die interne Validität und einer vor allem in älteren Psychotherapiestudien vorliegenden niedrigen Qualität, nehmen einige Autoren an, dass die gegenwärtige Datenlage die Wirksamkeit von psychotherapeutischen Interventionen überschätzt [11, 13]. Als Qualitätsmerkmale von RCT im Allgemeinen und Psychotherapiestudien im Speziellen werden Kriterien diskutiert wie eine operationalisierte Diagnosesicherung, hohe Qualifikation der Therapeuten, Monitoring der Therapietreue und Supervision der Therapeuten in allen Bedingungen, Analyse aller Daten der initial randomisierten Probanden (*intention-to-treat* (ITT) Analyse) und Manualisierung aller Bedingungen [11, 13, 14]. In einer methodenfokussierten Metaanalyse konnte gezeigt werden, dass die Effektstärken für Psychotherapieverfahren bei depressiven Störungen jeweils signifikant kleiner ausfielen in Studien, in denen die Kontrollbehandlungen manualisiert waren, die in dieser Bedingung tätigen Therapeuten geschult und supervidiert wurden, sowie deren Therapietreue überwacht wurde [14].

Ein weiterer Einflussfaktor auf die Effektstärke der Intervention ist die Art der Kontrollbedingung. Die Wahl der Kontrollbedingung kann unter anderem von der Phase der Therapieentwicklung abhängen, wobei in frühen Phasen der Fehler zweiter Art minimiert werden

sollte (d.h. einen vorhandenen Effekt nicht zu finden und somit die Weiterentwicklung einer wirksamen Therapie zu verhindern). Im Verlauf sollte der Fehler erster Art minimiert werden (d.h. einen Effekt zu finden, der nicht vorhanden ist und damit eine unwirksame Therapie zu etablieren) [14].

In Psychotherapiestudien lassen sich folgende Arten von Kontrollbedingungen unterscheiden [12, 14],

- Kontrollbedingungen ohne jegliche Behandlung.
- Wartelisten: Patienten auf einer Warteliste werden zu einem späteren Zeitpunkt in die Behandlung aufgenommen. Im Wartelistenvergleich fallen die Effektstärken meist höher aus als bei anderen Kontrollbedingungen [14, 15, 16]. Hieraus schlussfolgerten einige Autoren, dass die Wartelistenbedingung wie ein *nocebo* wirke, da sie verhindere, dass Patienten Ressourcen aktivieren und alternative Hilfsangebote suchen [12, 15].
- Placebotabletten.
- Kontrollbedingungen, die durch die Studie definiert werden, *ohne* spezifische Behandlungskomponenten. Hierbei wird Kontakt- und Aufmerksamkeitsintensität in der Kontrollgruppe und Interventionsgruppe gleich gehalten, ohne dass die Kontrollgruppe die spezifische Intervention erhält. Diese Art der Kontrollbedingung fördert eine hohe interne Validität.
- Kontrollbedingungen, die durch die Studie definiert werden, *mit* spezifischen Behandlungskomponenten, das heißt Teile der Intervention werden auch in der Kontrollbedingung angewendet, andere weggelassen (z.B. in *dismantling* Studien).
- Kontrollbedingungen, die nicht durch die Studie definiert sind wie *treatment as usual* (TAU) oder *care as usual*. Das heißt, die Behandlung, die die Patienten üblicherweise erhalten würden. Damit können Arzt- und Personalgespräche zur Diagnostik und Medikationseinstellung gemeint sein. Im stationären Kontext eventuell zusätzlich Psychoedukation und komplementäre Therapien. Charakteristisch für TAU ist, dass das TAU - Prozedere nicht zu Studienzwecken modifiziert oder strukturiert wird. Diese Art der Kontrolle besitzt eine hohe externe Validität, ist jedoch schwer zwischen Studien vergleichbar.
- Aktive Vergleichsgruppen: Die Patienten erhalten eine bereits etablierte oder alternative Behandlungsform gleicher Modalität (z.B. KVT versus CBASP). Diese Art des Vergleichs kommt eher in einer späteren Phase der klinischen Prüfung zur Anwendung.

1.2.4 Wirksamkeit von Psychotherapie im ambulanten und stationären Rahmen

Die meisten Nachweise aus Wirksamkeitsstudien liegen aus dem ambulanten Bereich vor. Hierzu wurden zahlreiche Metaanalysen veröffentlicht, die verschiedene Aspekte differenzierten. Die meisten Studien liegen zur KVT vor. Die Wirksamkeit von ambulanter KVT zeigte in neueren Metaanalysen eine mittlere Effektstärke von $g = 0.71 - 0.75$ [15, 17]. KVT zeigte sich im Vergleich jedoch weder anderen Therapieformen, noch reiner Pharmakotherapie überlegen, wobei die Kombination aus KVT und Pharmakotherapie der reinen Pharmakotherapie überlegen war [17]. Eine Metaanalyse von individuellen Patientendaten fand keinen Einfluss der initialen Depressionsschwere auf die Wirksamkeit der KVT, anders als es für die Pharmakotherapie beschrieben wurde [18]. Bei der Synthese von Effekten weiterer Psychotherapieformen depressiver Störungen (KVT, IPT, „dritte Welle“-Therapien, PDT u.a.) zeigt sich eine ähnlich hohe mittlere Effektstärke von $g = 0.70$ (95% CI: 0.64 – 0.75, $k = 369$) [16]. Deutlich wird hier das quantitative Gefälle zwischen Studien der verschiedenen Verfahren, denn in diese Metaanalyse flossen 192 Vergleiche zu KVT, zu PDT jedoch nur 11 ein. Zwischen den Effektstärken für die einzelnen Verfahren bestand kein signifikanter Unterschied [16]. In einer Netzwerkmetaanalyse, die Verfahren direkt miteinander verglich (KVT, BA, IPT, PDT u.a.) zeigten sich ebenfalls keine signifikanten Unterschiede in der Wirksamkeit zwischen den Therapieverfahren, mit der Ausnahme einer kleinen Überlegenheit von IPT gegenüber supportiver Psychotherapie [19]. In dieser Analyse zeigten sich Interventionen kürzerer Dauer denen längerer Dauer überlegen [19].

Adjustiert für Studienqualität, Art der Kontrollbedingung und Publikationsbias fallen die Effektstärken regelhaft kleiner aus [13, 15, 17]. So waren in einer der neueren, oben zitierten Metaanalysen lediglich 22% der Studien mit niedrigem Risiko einer systematischen Verzerrung behaftet; die in dieser Subgruppe ermittelte Effektstärke der Intervention lag nur noch bei $g = 0.38$ (95% CI: 0.32 – 0.44, $k = 71$). Adjustierung für einen Publikationsbias ergab $g = 0.31$ (95% CI: 0.24 – 0.38, $k = 84$) [16].

Zu stationärer Psychotherapie liegt eine Metaanalyse aus dem Jahre 2011 vor [20]. Diese Metaanalyse konnte in 15 Vergleichen ($n = 570$) einen kleinen positiven Effekt von $g = 0.29$ (95% CI: 0.13 – 0.44) der stationären Psychotherapie gegenüber den Kontrollbedingungen zeigen. In dieser Metaanalyse wurden jedoch einige Studien eingeschlossen, in denen auch die Kontrollgruppe mit Psychotherapie behandelt wurde und die Interventionsgruppe lediglich eine zusätzliche oder speziellere Intervention erhielt.

1.3 Fragestellung und Hypothesen

Zusammenfassend lässt sich sagen, dass nur eine Minderheit der depressiven Patienten eine stationäre Behandlung in Anspruch nimmt. Diese Minderheit ist häufig durch eine besonders schwere Depression, eine hohe Anzahl psychiatrischer Komorbiditäten oder Suizidalität charakterisiert oder hat bereits erfolglos ambulante Behandlungsversuche unternommen. Das Angebot stationärer Psychotherapie erfordert seitens der Klinik die koordinierte Umsetzung eines Therapiekonzepts, das Vorhalten ausgebildeter Therapeuten, sowie die Bereitstellung von Supervisoren. In Anbetracht der erforderlichen Ressourcen ist es daher notwendig zu evaluieren, ob stationär behandelte Patienten von einem strukturierten Psychotherapieangebot während einer kostspieligen Krankenhausbehandlung profitieren. Das Anliegen dieser Metaanalyse ist daher, den Effekt stationärer Psychotherapie auf die Depressionsschwere bei Patienten mit einer unipolar depressiven Störung aktualisiert zu schätzen. Hierbei sollten folgende Hypothesen (H) untersucht werden:

- H1: Stationäre Psychotherapie unipolar depressiver Störungen ist einer Standardbehandlung in der Reduktion der Depressionsschwere überlegen.
- H2: Die Effektstärke der psychotherapeutischen Intervention ist umso geringer, je standardisierter die Kontrollbedingung gestaltet ist.
- H3: Die Effektstärke ist abhängig von der Anzahl der Therapieeinheiten.
- H4: Die Effektstärke der Intervention unterscheidet sich zwischen Selbst- und Fremdbeurteilungsinstrumenten.

Kapitel 2

Methodik

2.1 Auswahl der Methode

Zur Beantwortung der Fragestellung, ob eine stationär angebotene Psychotherapie bei unipolar depressiven Patienten gegenüber Kontrollbedingungen ohne Psychotherapie überlegen ist, wurde die Methode der systematischen Literaturrecherche und Metaanalyse gewählt. Durch eine Metaanalyse wird die Präzision einer Effektschätzung durch Zusammenfassung aller Studien, die den interessierenden Effekt in der Population untersuchen, erhöht. Als Nullhypothese liegt die Annahme keines Unterschieds zwischen Intervention und Kontrollbedingung zugrunde. Das Ergebnis einer Metaanalyse unter diesen Voraussetzungen ist eine standardisierte Effektstärke, die eine Aussage und Vorhersage darüber zulässt, wie groß die Wirkung der interessierenden Intervention in der Population ist [21, S. 9-13].

2.2 Festlegung der Ein- und Ausschlusskriterien

Es wurden nur Studien in Betracht gezogen, deren Design einen Vergleich einer Interventionsgruppe mit einer Kontrollgruppe vorsah. Hierbei galt als Intervention ein Psychotherapieverfahren, das in manualisierter Form im stationären Rahmen durchgeführt werden kann. Ferner galten die Einschlusskriterien: Diagnose einer *major depressive episode* oder *disorder* (MDD) bzw. einer unipolar depressiven Episode oder Störung gemäß DSM¹- oder ICD-10 Kriterien, Ergebnisse einer Studie sollten in einer Form vorliegen, welche die Verwendung in einer Metaanalyse ermöglicht, Verwendung eines gängigen psychometrischen Instruments zur Erfassung der Depressionsschwere. Hierbei erhielten Selbstbeurteilungsskalen den Vorrang (BDI, QIDS-SR) in der Auswertung, da diese nicht für einen *detection bias* anfällig sind (siehe 2.6). Nachrangig waren Fremdbeurteilungsmaße, was bedeutet, dass in die Hauptanalyse nur ein Ergebnis einer Fremdbeurteilung eingeschlossen wurde, wenn keine Selbsteurteilung vorlag.

¹Versionen -III bis -5.

2.3 Datenbankrecherche und Studienselektion

Vier große biomedizinischen Datenbanken (CENTRAL, MEDLINE, PsycINFO und Web of Science) wurden nach relevanten Einträgen unter Anwendung einer replizierbaren Suchstrategie durchsucht [22, Appendix S1, siehe auch S. 323]. Die Ergebnisse der Datenbanksuche wurden in eine EndNote-Bibliothek importiert unter automatischem und manuellem Ausschluss von Duplikaten. Zwei Studienautoren selektierten relevante Artikel unabhängig von einander. Um die Auswahl zu vereinfachen, wurden mit Hilfe von EndNote *smart groups* Titel ausgewählt, die Schlagwörter enthielten, die einen Studienausschluss nahelegten. So wurden beispielsweise alle Titel herausgefiltert, die die Wörter 'childhood' 'adolescent', 'child' oder irrelevante Diagnosenkategorien wie 'bipolar', 'borderline', 'schizophrenia' im Titel trugen. Die aussortierten Titel wurden daraufhin noch einmal manuell auf Relevanz durchsucht. Nach der Reduktion der initialen Suchresultate auf 158 Titel, wurden diese im Folgenden manuell anhand ihres Volltextes eingeschlossen bzw. aussortiert. Das Ergebnis dieses Prozesses der Studienselektion ist im Appendix S2 der Originalpublikation in Form eines Flussdiagramms dargestellt [22, Appendix S2, Figure S1]. Diskrepanzen zwischen den Autoren in der initialen Studienselektion ließen sich durch gemeinsame Diskussion lösen.

2.4 Datenextrahierung

Die Datenextrahierung beinhaltete die tabellarische Zusammenfassung der relevanten Studieninformationen. Hierbei wurden Informationen über Stichproben- und Studiencharakteristika zusammengefasst wie Stichprobengröße, Geschlechterverteilung, Anzahl der Studienabbrecher, Dauer der Studie, Ein- und Ausschlusskriterien, publizierte Protokolle, Interventionsart und -frequenz, Kontrollbedingung, verwendete Manuale sowie Mittelwerte und Standardabweichungen der Studienergebnisse.

Die Art der Kontrollbedingung wurde im Hinblick auf die spätere Subgruppenanalyse nach folgenden Gesichtspunkten klassifiziert [angelehnt an 12]:

- a) Keine Behandlung oder Wartelistenbedingung,
- b) unstandardisierte TAU: Bedingungen, die nicht manualisiert waren und nicht durch die Studie vorgegeben wurden. Hierzu gehörten Aufenthalt im stationären Milieu, Pharmakotherapie, Arztkontakte und komplementäre Therapieangebote (Ergotherapie, Musik- oder Kunsttherapie).
- c) Standardisierte Kontrollbedingungen: Manualisierte oder sonstig strukturierte Kontrollbehandlung, die in Intensität oder Kontaktfrequenz der Intervention ungefähr äquivalent war, jedoch ohne die spezifischen Interventionsmerkmale.

Für die *risk of bias* - Analyse wurden Zitate extrahiert, die zur Beurteilung der Domänen des *Cochrane Collaboration's Tool for Assessing Risk of Bias* [23] herangezogen wurden.

2.5 Metaanalyse

Für die Hauptanalyse wurden die kontinuierlichen Daten der Selbstbewertungsskalen der Depressionsschwere herangezogen. Das Grundprinzip einer Metaanalyse über kontinuierliche Daten besteht darin, einen zusammenfassenden Schätzer des Interventionseffektes (im folgenden Summeneffektstärke) und dessen Konfidenzintervall anhand der gewichteten Einzeleffektstärken der eingeschlossenen Studien zu berechnen. Hierfür wurde in diesem Fall ein Modell herangezogen, das zudem die Varianz zwischen den Studien berücksichtigt (*random effects model*). In Subgruppenanalysen wurde der Effekt moderierender Variablen untersucht. Anhand eines *funnel plots* und Rangkorrelationstests wurde ein potentieller Publikationsbias untersucht. Im Folgenden werden die grundlegenden Formeln und Prinzipien dargestellt, die dem Verfahren zugrunde liegen.

2.5.1 Standardisierte Mittelwertsdifferenz

Für eine Metaanalyse über kontinuierliche Daten wird eine standardisierte, d.h. vom Messinstrument unabhängige Effektstärke für jede Einzelstudie berechnet. Der Begriff Einzelstudie kann hier synonym zu Studie oder Vergleich verstanden werden. Im folgenden soll für jeden Vergleich, der in eine Metaanalyse eingeht der Begriff Einzelstudie verwendet werden.

Die Effektstärke einer Einzelstudie geht mit dem Gewicht ihrer inversen Varianz in eine Metaanalyse ein. Hinreichend zur Berechnung sind diesem Fall die Differenz der Mittelwerte der unabhängigen Interventions- bzw. Kontrollgruppen ($\hat{\mu}_e - \hat{\mu}_c$), die jeweiligen Standardabweichungen (s_e, s_c) sowie die Stichprobengrößen (n_e, n_c). Das verwendete R-Paket `meta` [24, 25, 26] berechnet die standardisierte Mittelwertsdifferenz (SMD) einer Einzelstudie k als *Hedge's g* [27] nach

$$\hat{g}_k = \left(1 - \frac{3}{4n_k - 9}\right) \frac{\hat{\mu}_{ek} - \hat{\mu}_{ck}}{\sqrt{((n_{ek} - 1)s_{ek}^2 + (n_{ck} - 1)s_{ck}^2) / (n_k - 2)}}. \quad (2.1)$$

Der zweite Term der Gleichung repräsentiert die standardisierte Mittelwertsdifferenz, wobei die Standardisierung durch Division durch die gepoolte Standardabweichung erreicht wird. Der erste Term ist ein Korrekturfaktor, da die Effektstärke in kleinen Studien ohne diese Korrektur überschätzt würde [27]. Die Effektstärke ist interpretierbar wie *Cohen's d*², wird durch den Korrekturfaktor jedoch zu *Hedge's g* konvertiert. Die Varianz von \hat{g}_k ist

²Mit 0.2 als kleinem, 0.5 als mittlerem und 0.8 als großem Effekt [28]

$$\widehat{\text{Var}}(\hat{g}_k) = \frac{n_k}{n_{ek} \cdot n_{ck}} + \frac{\hat{g}_k^2}{2(n_k - 3.94)}. \quad (2.2)$$

Aus der Quadratwurzel der Varianz wird der Standardfehler ermittelt, anhand dessen das 95%-Konfidenzintervall des Schätzers \hat{g}_k festgelegt wird [26, S. 25-26].

2.5.2 Random effects model

Die SMD der Einzelstudien werden zur Schätzung der Summeneffektstärke zusammengefasst unter der Nullhypothese, dass die Summeneffektstärke sich nicht signifikant von Null unterscheidet. Die Schätzung basiert entweder auf einem *fixed effect model* (FEM) oder einem *random effects models* (REM) [21, S. 77-86]. Beide Modelle basieren auf der Berechnung des gewichteten Mittelwerts und der Varianz der SMD. Unterschiede zwischen FEM und REM bestehen dabei in der Gewichtung der Einzelstudien. Die Einzelstudien gehen jeweils mit dem Gewicht ihrer inversen Varianz (Innerstudienvarianz) in den Mittelwert ein, um die Präzision der einzelnen Messungen abzubilden. Je höher die Varianz, desto geringer die Präzision und desto geringer das Gewicht der Studie im Gesamtmodell. Im FEM wird davon ausgegangen, dass es in der Population *eine* wahre Effektstärke gibt, die allen beobachteten Einzelstudien zugrunde liegt und von diesen unterschiedlich präzise gemessen wurde [21, S. 59]. Das REM geht dagegen von der Grundannahme aus, dass es eine Verteilung von wahren Effektstärken gibt, die normal um einen Mittelwert streuen. Die Effekte der Einzelstudien im REM sind Realisationen dieser Verteilung. Sie variieren nicht nur durch unpräzise Messung, sondern auch weil die wahren Effekte aufgrund unterschiedlicher Populationen und Studiencharakteristika variieren [21, S. 69-71].

Konsequenz aus dieser Grundannahme des REM ist, dass in die Gewichtung der einzelnen Effektstärken im metaanalytischen Modell nicht nur die Innerstudienvarianz, sondern auch die Zwischenstudienvarianz (τ^2) eingehen muss. Die Zwischenstudienvarianz ist die Abweichung der Einzeleffektstärken von ihrem Mittelwert. Die Wahl des REM ggü. dem FEM in unserem Fall liegt in der theoretischen Überlegung begründet, dass die eingeschlossenen Studien unterschiedliche psychotherapeutische Verfahren anwenden (KVT, IPT, PDT etc.), die unterschiedlich effektiv sind und denen keine einzelne wahre Effektstärke zugrunde liegen kann. Im Folgenden soll formal nur das REM ausgeführt werden.

Die Summeneffektstärke $\hat{\theta}_R^3$, das heißt, das das Ergebnis einer Metaanalyse, ist wie oben beschrieben der gewichtete Mittelwert der Effektstärken der Einzelstudien nach

$$\hat{\theta}_R = \frac{\sum_{k=1}^K w_k^* \hat{\theta}_k}{\sum_{k=1}^K w_k^*}, \quad (2.3)$$

³Der verwendete Index R steht hier für *random effects model*.

wobei $\hat{\theta}_k = \hat{g}_k$. Hierbei ist das Gewicht w_k^* die inverse Varianz $\hat{\sigma}_k^2$ der Einzelstudie plus der Zwischenstudienvarianz τ^2 , $w_k^* = 1/(\hat{\sigma}_k^2 + \tau^2)$. Die Zwischenstudienvarianz τ^2 lässt sich auf unterschiedliche Weise schätzen [26, S. 36]. Hier wurde die Methode nach DerSimonian und Laird [29] verwendet. Die Varianz der Summeneffektstärke $\widehat{\text{Var}}(\hat{\theta}_R)$ ist die inverse Summe der Einzelstudiengewichte nach

$$\widehat{\text{Var}}(\hat{\theta}_R) = \frac{1}{\sum_{k=1}^K w_k^*}. \quad (2.4)$$

Der Standardfehler lässt sich als Quadratwurzel der Varianz ermitteln und somit das Konfidenzintervall um $\hat{\theta}_R$ festlegen. Wenn die Zwischenvarianz $\tau^2 = 0$ beträgt, sind die Summeneffektschätzer des REM und FEM identisch.

2.5.3 Heterogenität

Durch die Zwischenstudienvarianz lässt sich zudem die Heterogenität der Studien quantifizieren und testen, ob die Einzeleffekte der Studien in der Metaanalyse mehr als erwartet variieren. Die Gesamtvarianz setzt sich zusammen aus Zufallsvarianz und tatsächlicher Heterogenität, die durch Unterschiede im Studiendesign oder den Stichproben bedingt ist. Die Heterogenität beschreibt, wie groß die Dispersion der Einzeleffektstärken in der Stichprobe ist. Zur ihrer Schätzung wird *Cochran's Q* berechnet [21, S. 109]: Die gewichtete Summe der quadratischen Abweichungen der Einzeleffektstärken von ihrem Mittelwert gemäß

$$Q = \sum_{k=1}^K w_k (\hat{\theta}_k - \hat{\theta}_F)^2. \quad (2.5)$$

Die Variable $\hat{\theta}_k$ repräsentiert hier global den Schätzer der Einzeleffektstärken, der in unserem Falle, wie in 2.4.1 beschrieben, gleich \hat{g}_k ist (siehe Formel 2.1). Das Symbol $\hat{\theta}_F$ wiederum steht für den Mittelwert der k Studien bzw. die Summeneffektstärke nach dem *fixed effects model*. Auf die Formel des letzteren soll hier nicht eingegangen werden, da sie Formel 2.3 entspricht, mit dem Unterschied, dass anders als w_k^* , w_k als $1/\hat{\sigma}_k^2$ definiert ist, die Zwischenstudienvarianz τ^2 also nicht einbezogen wird. Anhand der Gesamtvarianz Q lässt sich durch die χ^2 -verteilte Testgröße $Q - df$ testen, ob die Varianz größer ist, als unter der Anzahl der Studien erwartet [21, S. 109-112]. Die Signifikanztestung von Q hat jedoch bei einer geringen Studienzahl eine niedrige *power* [30]. Zur besseren Quantifizierung der Heterogenität wird daher der Index I^2 angegeben [30, siehe 2.6]. Dieser gibt auf einer Skala von 0-100% an, wie groß der Anteil der Heterogenität (Zwischenstudienvarianz) an der Gesamtvarianz ist:

$$I^2 = \left(\frac{Q - df}{Q} \right) \times 100\% \quad (2.6)$$

I^2 ist ein intuitiv verständliches Maß der Konsistenz einer Metaanalyse und kann zwischen Metaanalysen verglichen werden, da es von der Zahl der eingeschlossenen Studien unabhängig ist [30]. Als Orientierung wurde vorgeschlagen 25% als geringe, 50% als mittlere und 75% als hohe Heterogenität zu klassifizieren [30].

Eine hohe Heterogenität der Studien deutet darauf hin, dass die Schlussfolgerungen der Metaanalyse eingeschränkt generalisierbar sind, da jede der Studien einen unterschiedlichen Effekt misst. Eine niedrige Heterogenität bedeutet, dass alle Studien in der Metaanalyse wahrscheinlich den gleichen Effekt messen und die beobachtete Varianz zwischen ihnen lediglich zufallsbedingt ist [21, S. 117-119]. Die Heterogenität einer Metaanalyse kann durch heterogene klinische oder Studiencharakteristika bedingt sein, ist aber nicht mit diesen gleich zu setzen. So können sich die Interventionen unterscheiden, ohne dass sich dies in der statistischen Heterogenität spiegelt.

2.5.4 Subgruppen- und Sensitivitätsanalysen

Wird die Heterogenität einer Metaanalyse als hoch eingeschätzt, können Studiencharakteristika herangezogen werden, die als moderierende Variablen Heterogenität teilweise aufklären können. Moderierende Variablen (Moderatoren) können in einer Metaregression als Prädiktoren oder in einer Subgruppenanalyse als Gruppierungsvariablen eingesetzt werden. Hierbei gilt, je höher die Anzahl untersuchter Moderatoren, desto wahrscheinlicher wird ein falsch-positiver Befund aufgrund der Alphafehler-Kumulierung [31]. Eine grobe Empfehlung ist, nicht mehr als einen Moderatoreffekt zu testen, wenn weniger als 10 Studien in der Metaanalyse eingeschlossen wurden [23, 26, S. 88]. Für die Berechnung der Subgruppeneffekte gibt es unterschiedliche Verfahren [21, Kap. 19]. Hier wurde ein *random effects model* mit einem Q -Test für Heterogenität verwendet, der eine gemeinsame Varianz in den Subgruppen annimmt [21, S. 178 ff.]. In dieser Metaanalyse wurden zu Beginn die Hypothesen aufgestellt, dass die Effektstärke mit *i*) der Anzahl der Therapiestunden, *ii*) mit der Art der Kontrollbedingungen und *iii*) mit der Art des *outcome* Erhebungsmodus (Selbst- vs. Fremdrating) variiert. Anhand dieser Hypothesen wurden mit kategorialen Variablen Subgruppenvergleiche durchgeführt.

2.5.5 Ermittlung des Publikationsbias

Die Prüfung auf Vorhandensein eines Publikationsbias basiert auf der Annahme, dass eine Studie mit höherer Wahrscheinlichkeit publiziert wird, wenn sie einen signifikanten Effekt ausweist. Damit Effekte in kleinen Stichproben – die den wahren Effekt unpräzise messen und bei Messwiederholungen deshalb stärker variieren – das festgelegte Signifikanzniveau

unterlaufen, müssen die Effektstärken umso größer sein. Es würde folglich durch die Tendenz zur bevorzugten Veröffentlichung signifikanter Effekte eine Selektion zugunsten kleiner Studien mit unpräzise gemessenen, zufällig großen Effektstärken stattfinden [32]. Metaanalysen, die die Effekte ausschließlich kleiner, veröffentlichter Studien zusammenfassen, würden den wahren Effekt folglich überschätzen. Theoretisch sollten ebenso viele Studien existieren, jedoch unveröffentlicht bleiben, die den Effekt als kleiner oder sogar gegenläufig schätzen [32].

Um zu bestimmen, ob eine Metaanalyse von einem Publikationsbias bzw. einem *small-study effect*⁴ betroffen ist, wurde zur grafischen Beurteilung ein konturverstärkter *funnel plot* erstellt und dessen Asymmetrie durch einen Rangkorrelationstest [33] auf Signifikanz geprüft.

2.6 Qualitätsbeurteilung

Verschiedene Werkzeuge zur Bewertung der Studienqualität können herangezogen werden. Das Cochrane Handbook [34] rät hierbei von Punktsystemen und quantitativen Maßen ab und schlägt das *Cochrane Collaboration's Tool for Assessing Risk of Bias* [23] vor, dessen Domänen empirisch validierte Risiken für die interne Validität identifizieren sollen. Die fünf Bewertungsdomänen beziehen sich auf Generierung einer Zufallssequenz und randomisierte Zuweisung der Teilnehmer auf die Bedingungen (*selection bias*), Verblindung der Teilnehmer und der Kliniker (*performance bias*), Verblindung der *outcome* Assessoren (*detection bias*), auf die Vollständigkeit der Darstellung der relevanten Ergebnisse (*reporting bias*) und die Vollständigkeit der Daten (*attrition bias*). Da Psychotherapiestudien die methodische Besonderheit inne ist, dass die Therapeuten und die Probanden nicht für die Behandlung verblindet werden können, entfiel in unserer Metaanalyse die Bewertung des *performance bias*. Für jede der übrigen vier Domänen wurden die Studien als mit geringem, hohem oder unklarem Risiko für einen *bias* bewertet.

- *Selection bias*: Alle Studien, die die Generierung der Randomisierungssequenz, sowie die verdeckte Allokation der Probanden auf die Bedingungen beschrieben, wurden mit geringem Risiko bewertet. Sofern die Studien nicht randomisiert waren, wurden das Risiko als hoch eingestuft. Ebenso als hoch risikobehaftet wurden Studien eingestuft, die die Behandlungsbedingungen in konsekutiver Reihenfolge anboten, wodurch die Zuweisung der Probanden zu einer Bedingung beim Einschluss in die Studie vorhersagbar war. Als unklar wurde das Risiko eingestuft, wenn die Stichprobe

⁴Der Begriff *small-study effects* bildet die deskriptive Überkategorie für das Phänomen, dass überwiegend kleine Studien mit großen Effekten verfügbar sind, was sich im *funnel plot* als Asymmetrie darstellt. Der Publikationsbias ist dabei nur ein Mechanismus, der zu einem *small-study effect* führen kann, wird jedoch am häufigsten diskutiert [25, S. 107]. Andere Ursachen für *small-study effects* können darin liegen, dass kleinere Studien aufgrund methodischer Mängel oder spezifischerer Stichproben mit geringerer Varianz größere Effekte schätzen. Nicht nur, weil sie aufgrund ihrer Signifikanz eher publiziert werden [32].

als randomisiert bezeichnet wurde, jedoch keine Details zu Sequenzgenerierung und Zuweisung der Probanden angegeben wurden.

- *Detection bias*: Sofern das klinische Personal für die Erfassung der Ergebnisvariablen unabhängig und verblindet war, wurde das Risiko für einen *detection bias* als gering bewertet. Bei fehlender Verblindung galt das Risiko als hoch, bei fehlenden Angaben zur Verblindung als unklar. Die Erfassung einer Selbstbewertung wurde ebenfalls als gering risikobehaftet bewertet.
- *Attrition bias*: Alle Studien, die eine ITT-Analyse durchgeführt haben, fehlende Daten imputiert haben, Angaben über Anzahl und Charakteristika fehlender Daten gemacht oder keine fehlenden Daten verzeichnet haben, wurden als gering risikobehaftet bewertet. Das Risiko wurde als hoch bewertet, wenn keine Angaben über die fehlenden Daten gemacht wurden und keine ITT-Analyse berichtet wurde.
- *Reporting bias*: Sofern ein vorab veröffentlichtes Studienprotokoll existierte, wurde das Risiko für *reporting bias* als gering eingeschätzt. Hoch, sofern erwartbare Angaben zu den Studienresultaten fehlten und unklar, wenn kein veröffentlichtes Studienprotokoll verfügbar war.

Kapitel 3

Ergebnisse

3.1 Ergebnisse der Hauptmetaanalyse

In der Hauptmetaanalyse der kontinuierlichen Selbstbeurteilungsdaten der Depressionsschwere zeigte sich ein kleiner Effekt von $g = 0.24$ (95% CI: 0.12 – 0.37, $k = 19$, $n = 1080$, $p < 0.001$). Der Effekt war homogen ($I^2 = 0\%$, $Q(df = 18) = 14.9$, $p = 0.67$), was bedeutet, dass es keine systematische Variation zwischen den Studien gibt und die Unterschiede in den Einzeleffektstärken eher auf Zufallsvariation zurück zu führen sind [22, Figure 1, S. 330].

Die Effektstärke entspricht einer Reduktion der Depressionsschwere um ca. 2.2 Punkte auf der BDI-Skala¹ und einer *number needed to treat* (NNT) von 7.4. Bei kontinuierlichen Daten ist die NNT nicht intuitiv interpretierbar. Sie gibt die Anzahl an Patienten an, die mit der Intervention behandelt werden muss, um bei einem zusätzlichen Patienten den gegenüber der Kontrollbedingung überlegenen Effekt zu erzielen, wenn die gleiche Anzahl an Patienten mit der Kontrollbedingung behandelt worden wäre [35].

In einer *leave-one-out* Analyse, bei der die Metaanalyse jeweils unter Ausschluss einer Einzelstudie wiederholt wird, zeigte sich kein dominierender Einfluss einer Einzelstudie.

Aus den *follow-up* Daten ging ebenfalls ein kleiner Effekt von $g = 0.21$ (95% CI: 0.01 – 0.41, $k = 9$, $n = 710$, $p = 0.04$) mit niedriger bis mittlerer aber nicht signifikanter Heterogenität ($I^2 = 29.8\%$, $Q(df = 8) = 11.39$, $p = 0.18$) hervor [22, Figure 3, S. 331]. Dieses Ergebnis spricht jedoch für eine Konstanz der Wirkung der stationären Psychotherapie über einen Zeitraum von 12 Monaten. Für die *follow-up* Daten gilt, dass sie unter naturalistischen Bedingungen erhoben wurden und eine hohe Anzahl an Patienten in den Einzelstudien keine *follow-up* Daten zur Verfügung gestellt haben (Median 30%) [22, S. 325].

3.1.1 Ergebnisse der Subgruppenanalyse

Für die Studien der Subgruppe mit niedriger Standardisierung der Kontrollbedingung (Warteliste, TAU) ergab sich eine Summeneffektstärke von $g = 0.35$ (95% CI: 0.17 – 0.53, $k = 10$,

¹Hierfür wurde die gepoolte Standardabweichung aller BDI-Daten der eingeschlossenen Studien zugrunde gelegt (SD 9.19).

$n = 560$, $p < 0.001$, $I^2 = 0.3\%$). Für Studien mit hoher Standardisierung der Kontrollbedingung ergab sich ein von Null nicht signifikant unterschiedlicher Effekt $g = 0.14$ (95% CI: $-0.03 - 0.32$, $k = 9$, $n = 520$, $p = 0.10$, $I^2 = 0.0\%$). Der Unterschied zwischen den Subgruppen war nicht statistisch signifikant ($Q(df = 1) = 2.67$, $p = 0.10$) [22, Figure 2, S. 331]. Dies deutet darauf hin, dass der positive Effekt der Intervention hauptsächlich durch Studien mit unstandardisierten Kontrollbedingungen bedingt ist.

Die Subgruppe der Studien, die Ergebnisse von Fremdbeurteilungsmaßen angab, zeigte ebenfalls eine kleine, signifikante Summeneffektstärke von $g = 0.3$ (95% CI: $0.16 - 0.44$, $k = 15$, $n = 870$, $p < 0.001$, $I^2 = 0.0\%$). Die Differenz zwischen den Fremd- und Selbstbeurteilungsergebnissen der Studien war nicht signifikant ($Q(df = 1) = 0.69$, $p = 0.40$). Daraus lässt sich ableiten, dass Fremdbeurteilungsmaße in dieser Stichprobe zu gleichwertigen Ergebnissen kommen, wie Selbstbeurteilungsmaße.

Es fanden sich keine signifikanten Unterschiede zwischen den Subgruppen unterschiedlicher Therapieeinheiten (siehe Tabelle 3.1; $Q(df = 2) = 0.44$, $p = 0.80$). Hierbei zeigte die Gruppe mit mehr als 21 Therapiesitzungen einen von Null nicht signifikant unterschiedlichen Effekt. In dieser Subgruppe befanden sich jedoch auch überwiegend Studien mit standardisierter Kontrollbedingung [für eine Übersicht über die Anzahl an Therapieeinheiten siehe 22, Table 1, S. 326-329]. Bei einer kleinen Anzahl an Studien innerhalb der Subgruppen und Kovarianz der Moderatoren kann aus dieser Analyse kein eindeutiger Zusammenhang zwischen der Anzahl an Therapiesitzungen und der Wirksamkeit geschlossen werden.

TABELLE 3.1: Subgruppenvergleich nach Anzahl der Therapiesitzungen

Therapiesitzungen	k	g	95%-CI	I^2	p -Wert
≤ 10	6	0.33	[0.02; 0.63]	0.0%	0.038
11 – 20	8	0.24	[0.09; 0.39]	21.6%	0.002
≥ 21	5	0.18	[-0.12; 0.48]	0.0%	0.236

3.1.2 Ergebnisse der Tests auf Publikationsbias

Die Prüfung von *small-study effects* bzw. eines Publikationsbias erfolgte grafisch und durch einen Rangkorrelationstest.

In einem *funnel plot* [22, Appendix S2, Figure S3] wurden die Einzeleffektstärken auf der x -Achse gegen ihren jeweiligen Standardfehler auf der y -Achse abgetragen. Der Standardfehler gilt hierbei als Maß der Präzision und ist unmittelbar abhängig von der Stichprobengröße (skaliert mit $n^{-1/2}$). Je größer die Stichprobe, desto kleiner der Standardfehler. Ein Datenpunkt repräsentiert eine Studie. Die Datenpunkte gruppieren sich um ihren Mittelwert (die Summeneffektstärke), der orthogonal zur x -Achse steht. Im Falle der Abwesenheit eines Publikationsbias, streuen die kleinen unpräzisen Effekte mit hohem Standardfehler

symmetrisch um den Mittelwert an der Basis des Dreiecks, und an der Spitze die sehr präzisen Effekte mit geringem Standardfehler. Eine asymmetrische Verteilung der Datenpunkte deutet auf einen *small-study effect* hin; es kann angenommen werden, dass unveröffentlichte oder nicht gefundene Studien existieren, die keinen oder einen gegenteiligen Effekt aufzeigen [32].

Die Konturverstärkung im *funnel plot* zeichnet die Verteilungsbereiche unter der Nullhypothese ein, in denen die Datenlage entweder für die Beibehaltung der Nullhypothese spräche (weiß) oder bei farbkodiertem Signifikanzniveau von $0.1 - 0.05$, $0.05 - 0.01$ und < 0.01 für die Alternativhypothese (dunkel-, mittel- bzw. hellgrau). Lügen vor allem kleine Studien in den signifikanten Bereichen (grau) und größere Studien in den Bereichen keiner Signifikanz (weiß), so spräche dies für einen Publikationsbias bzw. *small study effect* [26, S. 112]. Innerhalb der grauen Konturen lagen in unserem Fall jedoch Studien jeder Größe, was gegen einen Publikationsbias spricht.

Die Verteilung der Datenpunkte in unserer Darstellung war symmetrisch innerhalb des Dreiecks um den Mittelwert [22, Appendix S2, Figure S3]. Die Nullhypothese der Symmetrie des *funnel plots* konnte durch Prüfung mit dem Rangkorrelationstest nicht zurückgewiesen werden ($z = 0.8, p = 0.42$), was bedeutet, dass keine *small-study effects* vorliegen, demzufolge auch kein Publikationsbias [33].

3.2 Ergebnisse der Qualitätsbeurteilung

Eine grafische Übersicht über die Ergebnisse der Qualitätsbeurteilung findet sich in den ergänzenden Darstellungen [22, Appendix S2, Figure S2]. Insgesamt lag eine gute Gesamtqualität der Studien vor. Die Diagnose wurde in allen bis auf eine Studie mit einem standardisierten Interview gesichert. Die Mehrzahl der Studien war randomisiert und durch den Einschluss von Selbstbeurteilungsergebnissen ist das Risiko eines *detection bias* insgesamt gering.

- *Selection bias*: Zwölf Studien beschrieben die Generierung der Randomisierungssequenz und Allokation der Probanden (niedriges Risiko). Fünf Studien waren nicht randomisiert (hohes Risiko) und zwei Studien machten Angaben aus denen das Vorgehen bei der Randomisierung nicht erschlossen werden konnte (unklares Risiko).
- *Detection bias*: Alle Studien berichteten Ergebnisse aus Selbstbeurteilungsverfahren, weshalb die Daten der Hauptmetaanalyse ein niedriges Risiko für *detection bias* haben. Dennoch bewerteten wir die fehlende Verblindung von Assessoren bei der Erhebung einer Fremdbeurteilung mit hohem Risiko. In einer Studie wurden die initial unverblindeten Beurteilungen durch eine nachträgliche verblindete Beurteilung von Videoaufzeichnungen validiert (unklares Risiko).

- *Attrition bias*: Vier Studien, die keine ITT-Analyse durchgeführt haben und keine Angaben zu Studienabbruchern einbezogen, wurden als hoch risikobehaftet eingestuft. Eine Studie erwähnte Studienabbrucher, doch es war nicht ersichtlich, ob diese vor oder nach der Randomisierung abgebrochen haben (unklares Risiko). Die übrigen Studien wurden mit einem geringen Risiko für *attrition bias* bewertet.
- *Reporting bias*: Drei Studien, die einem zuvor veröffentlichten Protokoll folgten bewerteten wir als gering risikobehaftet. Eine Studie wurde als hoch risikobehaftet bewertet, da Daten von zwei Interventionsgruppen nicht separat berichtet wurden. Alle anderen Studien, für die kein Protokoll verfügbar war, wurden mit unklarem Risiko bewertet.

Kapitel 4

Diskussion

4.1 Bedeutung der Ergebnisse

In dieser systematischen Übersicht und Metaanalyse zeigte sich ein kleiner, jedoch robuster Effekt von stationärer Psychotherapie auf die Depressionsschwere in der Akuttherapie ($g = 0.24$). Der Effekt zeigte sich in einer Subgruppe von Studien mit naturalistischen *follow-up*-Daten über einen Zeitraum von mindestens einem Jahr stabil ($g = 0.21$). Hierbei liegen jedoch keine systematischen Daten über Rezidive vor und es besteht ein hohes Risiko für einen *attrition bias* in Anbetracht fehlenden Rücklaufs.

Alle Probanden der eingeschlossenen Studien (mit Ausnahme einer Studie mit Wartelistenbedingung) haben eine stationäre Behandlung, mehrheitlich mit Pharmakotherapie, erfahren. Dadurch kam mutmaßlich allen Patienten der unspezifische Effekt zugute, der durch die bloße Teilnahme an einer Behandlung und der damit verbundenen Veränderungserwartung bedingt wurde. In der von uns ermittelten Effektstärke bildet sich daher der zusätzliche Nutzen stationärer Psychotherapie ab, der über diesen unspezifischen Effekt hinausgeht. Die Größe der Effektstärke ist vergleichbar mit früheren Ergebnissen zu stationärer Psychotherapie [20].

Der Effekt war zwar statistisch signifikant, ist jedoch von kleiner klinischer Relevanz, legt man eine *minimal clinical important difference* von 3 Punkten im BDI als Maßstab an [36]. Ein gleichermaßen klinisch grenzwertig relevanter Effekt ergab sich jedoch im Durchschnitt für Antidepressiva [6]. Die Effektstärke könnte gegenüber den diskutierten Effektstärken von 0.70–0.75 in ambulanten Stichproben [17, 16] aus verschiedenen Gründen geringer ausgefallen sein. Zum einen könnte durch das stationäre Setting eine größere Angleichung der Bedingungen stattgefunden haben, wodurch sich die Varianz zwischen Interventions- und Kontrollgruppen reduziert hat. Zum anderen könnten die stationär behandelten Patienten höhere Raten an Behandlungsresistenz, psychiatrischer Komorbidität oder eine schwerere Symptombelastung gezeigt haben. Es konnte bereits in der Vergangenheit gezeigt werden, dass Effektstärken kleiner ausfallen, wenn Probanden aus einer klinischen Population versus der Allgemeinbevölkerung rekrutiert werden [16]. Einige Studien in unserer Metaanalyse wiesen einen hohen Anteil von Patienten mit Komorbiditäten und therapieresistenter

Depression auf, andere Studien wiederum schlossen Patienten mit Therapieresistenz oder sehr schweren Depressionen aus [22, S. 325 - 329]. Insgesamt lässt sich anhand unserer Daten kein Zusammenhang zwischen Therapieresponse und Patientencharakteristika wie Depressionsschwere oder Komorbiditäten untersuchen; dies ließe sich in einer Metaanalyse mit individuellen Patientendaten realisieren. Ein weiterer Grund für die im Vergleich geringere Effektstärke besteht darin, dass die oben zitierten Metaanalysen eine hohe Anzahl an Wartelistenvergleichen einbezogen haben.

In unseren Subgruppenvergleichen zeigte sich, dass der positive Effekt der Intervention hauptsächlich durch Studien ohne standardisierte Kontrollbedingung getrieben wurde. In Studien mit manualisierter, balancierter Kontrollbedingung zeigte sich keine Überlegenheit der psychotherapeutischen Intervention. Dies bedeutet, dass sobald die Kontrollbedingung mit einer strukturierten, ähnlich frequenten aber unspezifischen Zuwendung behandelt wurde, die Patienten genauso stark profitierten wie die Patienten, die Psychotherapie erhielten. Das spricht zum einen dafür, dass die stationäre Implementierung von Psychotherapie sinnvoll ist, da psychotherapeutische Interventionen oder eine äquivalente Zuwendung einen robusten positiven Effekt über stationäre TAU hinaus haben. Dies bedeutet zum anderen jedoch, dass wir nicht wissen, welches die effektiven Mechanismen der Therapie sind. Es könnte sein, dass die strukturierte, geplante Zuwendung von Personal hinreichend ist, ohne dass eine theoriebasierte Intervention notwendig ist. Gleichzeitig kann auch ein von uns nicht berücksichtigter, unbekannter dritter Faktor, wie z.B. strengere Methodik oder eine höhere Anzahl an Sitzungen in dieser Subgruppe für das Ergebnis keines Unterschieds verantwortlich sein. Die Unterschiede zwischen den Subgruppen waren zudem nicht statistisch signifikant, weshalb aus den Ergebnissen lediglich ein Hinweis abgeleitet werden kann. Für die Zukunft der klinischen Forschung ergibt sich daher, dass in *dismantling* Studien die einzelnen Faktoren einer komplexen Therapie auf ihren Beitrag zur Wirksamkeit untersucht werden sollten. Auf Patientenseite sollten die Faktoren untersucht werden, die die Patienten für spezifische Interventionen qualifizieren (z.B. emotionaler Missbrauch in der Kindheit bei chronischen Depressionen für CBASP [37]).

Zu ähnlichen Ergebnissen bezüglich des Einflusses der Kontrollbedingung auf die Effektstärke kam eine Metaanalyse von Studien zu ambulanter und stationärer Psychotherapie depressiver Störungen [14]. Hierbei generierten Vergleiche mit Wartelisten und unstandardisierten TAU - Bedingungen ebenfalls die größten Effektstärken; im Vergleich mit aktiven Kontrollgruppen oder Placebotabletten fanden sich keine Effekte [14]. Bezogen auf die in der Einleitung erwähnten Entwicklungsphasen einer Therapie und die Arten von Kontrollbedingungen (siehe 1.2.3) diskutierten die Autoren dieser Metaanalyse, dass es keine 'richtigen' oder 'falschen' Kontrolldesigns gebe. Vielmehr, dass in den frühen Stadien der Therapieentwicklung im Vordergrund stehen sollte, eine ausreichende *power* zu erreichen, möglicherweise zu Lasten der internen Validität durch geringere Kontrolle und Standardisierung der Bedingungen (Vergleich einer größeren Stichprobe gegen z.B. Warteliste oder

Kontrollgruppe ohne Behandlung). Ein positiver Effekt aus dieser frühen Phase sollte dann unter stärker kontrollierten, randomisierten Bedingungen in einer Studie mit hoher interner Validität, auch *efficacy* Studie genannt, geprüft werden (z.B. Vergleich gegen standardisierte Kontrollbedingung mit unspezifischen Faktoren) [14]. Spezifische Therapiekomponenten könnten zudem anschließend in *dismantling* Studien auf ihren Beitrag zur Wirksamkeit geprüft werden (Vergleich gegen Kontrollgruppe mit selektierten spezifischen Wirksamkeitsfaktoren). In den letzten Stadien interessieren die Generalisierbarkeit der Ergebnisse im praktischen Versorgungskontext. Hierzu dienen größere Stichproben und naturalistische Bedingungen (Vergleich gegen unstandardisierte TAU oder gegen etablierte aktive Vergleichstherapien) [14].

Es zeigten sich keine signifikanten Unterschiede zwischen Studien unterschiedlicher Anzahl an Therapiesitzungen. Die Zusammensetzung der Gruppe mit ≥ 21 Sitzungen glich zum großen Teil der Subgruppe mit standardisierter Kontrollbedingung, weshalb ein unabhängiger Beitrag der Anzahl an Sitzungen nicht geschätzt werden kann. Hieran lässt sich auch illustrieren, dass bei einer kleinen Anzahl eingeschlossener Studien die Durchführung von mehr als einer Subgruppenanalyse nicht sinnvoll ist [23].

Zusammenfassend bedeuten unsere Ergebnisse für die Praxis der stationären Umsetzung von Psychotherapie, dass die Implementierung eines manualisierten Verfahrens einen positiven Effekt auf die Depressionsschwere über den bloßen stationären Aufenthalt hinaus bedeutet. Auch auf längere Sicht profitieren die Patienten von der stationären Psychotherapie. Spezifische Wirksamkeitsfaktoren gilt es in Zukunft zu spezifizieren. KVT besitzt die breiteste Evidenzbasis und ist auch in dieser Metaanalyse die am häufigsten untersuchte Methode. Die Förderung von Studien insbesondere zu stationären psychodynamischen Therapien ist wünschenswert, da diese mit nur einer Studie in dieser Metaanalyse vertreten sind.

4.2 Qualitätsaspekte und Methodik

In unserer Metaanalyse zeigte sich keine Heterogenität der Stichprobe, was darauf hinweist, dass die Studien den gleichen Effekt messen und die Schätzung der Summeneffektstärke dadurch an Reliabilität gewinnt. Für die Reliabilität der Schätzung spricht auch, dass sich nur ein geringer, nicht signifikanter Unterschied zwischen Selbst- und Fremdbeurteilungsergebnissen fand.

Ferner fand sich kein Hinweis auf einen Publikationsbias. Die Anzahl an Studien, die unsere Kriterien erfüllte war zwar klein, die Gesamtzahl an eingeschlossenen Patienten lag jedoch bei über 1000. Die Summeneffektstärke entsprach der Einzeleffektstärke einer der

eingeschlossenen Studien mit über 200 Probanden [38]. Das *risk of bias* war insgesamt niedrig, wobei fünf Studien keine oder eine unzureichende Randomisierung durchgeführt haben. Weitere Qualitätsmerkmale der eingeschlossenen Studien, die sich aus den Einschlusskriterien ergaben, war das Vorliegen operationalisierter Diagnosen und manualisierter Interventionen. Beide letztgenannten Merkmale wurden in vergangenen Metaanalysen mit einer konservativeren Schätzung der Effekte assoziiert [14, 16]. Zusammenfassend ergibt sich dadurch ein reliabler Schätzer der Effektstärke. Die Berechnung der NNT folgte in dieser Metaanalyse dem Verfahren von Kraemer und Kupfer [39]. In einem empirischen Vergleich konnte gezeigt werden, dass dieses Verfahren in einer geringeren Anzahl von Fällen die tatsächliche NNT schätzt, als eine neuere Methode [35]. Letztere Methode setzt jedoch voraus, dass die Responserate der Kontrollgruppe bekannt ist, die wir jedoch nicht ausgewertet haben [35].

4.3 Limitationen und Ausblick

Über die in der Originalpublikation diskutierten Punkte hinaus gilt anzumerken, dass diese Metaanalyse keine systematische Untersuchung von Nebenwirkungen stationärer Psychotherapie enthielt. Ferner wurden die Response- und Remissionsdaten nicht ausgewertet. Eine Analyse einer Verzerrung durch *investigator allegiance* wurde nicht durchgeführt. Die verfügbaren Daten ließen außerdem keine Aussage darüber zu, ob Patienten der eingeschlossenen Studien repräsentativ sind für stationär behandelte Patienten mit besonders schwerer depressiver Symptomatik, therapieresistenten depressiven Störungen, mit hohem Suizidrisiko und hoher psychiatrischer Komorbidität. Ebenfalls keine Auswertung erfuhren die Faktoren der Therapeutenqualifikation, Supervision oder Therapieadhärenz in beiden Bedingungen.

Desweiteren bleiben ökonomische Fragestellungen offen, bei denen interessiert, ob die stationäre Psychotherapie auch dazu beitragen kann, Kosten für das Gesundheitswesen insgesamt zu senken. Wenn ein großer Teil der Kosten, die durch depressive Störungen verursacht werden, auf die stationäre Behandlung entfallen [7], wäre wichtig zu erforschen, wie die stationäre Behandlung so gestaltet werden kann, dass im Nachgang direkte und indirekte Kosten eingespart werden können. Zukünftige Vergleiche gegen eine Wartelistenbedingung sollten vermieden werden, da dies die Ergebnisse nicht nur verzerrt, sondern auch bei bestehender Indikation für eine stationäre Behandlung einer depressiven Störung ethisch nicht vertretbar ist. Da stationäre psychotherapeutische Interventionen bei Depressionen bereits sehr verbreitet sind, gilt es in Zukunft, die Überlegenheit oder den Zusatznutzen eines Verfahrens gegen z.B. den aktiven Vergleich mit stationärer KVT zu suchen.

Literatur

- [1] Kessler RC und Bromet EJ. „The epidemiology of depression across cultures“. In: *Annual Review of Public Health* 34 (2013), S. 119–138.
- [2] Busch M, Maske U, Ryl L, Schlack R und Hapke U. „Prävalenz von depressiver Symptomatik und diagnostizierter Depression bei Erwachsenen in Deutschland“. In: *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 56.5-6 (2013), S. 733–739.
- [3] Hasin DS, Sarvet AL, Meyers JL, Saha TD, Ruan WJ, Stohl M und Grant BF. „Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States“. In: *JAMA Psychiatry* 75.4 (2018), S. 336–346.
- [4] Mathers CD und Loncar D. „Projections of global mortality and burden of disease from 2002 to 2030“. In: *PLoS Medicine* 3.11 (2006), e442.
- [5] DGPPN, BÄK, KBV und AWMF. *für die Leitliniengruppe Unipolare Depression*(2015) S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare Depression–Langfassung, 2. Aufl, Vers 2, Nov 2015. (* Organisationen, die in der Leitliniengruppe kooperierten: DGPPN, BÄK, KBV, AWMF, ACKPA, AkdÄ, BpTK, BApK, DAGSHG, DEGAM, DGPM, DGPs, DGRW, BDK, BDP, BPM, BVDN, BVDP, BVVP, CPKA, DÄVT, DFT, DGGPP, DGPT, DGVT, DPG, DPV, DPtV, DVT, GwG, Stiftung Deutsche Depressionshilfe)*. doi: 10.6101/AZQ/000266. 2016.
- [6] Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, Leucht S, Ruhe HG, Turner EH, Higgins JP et al. „Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis“. In: *Lancet* 391.10128 (7. Apr. 2018), S. 1357–1366.
- [7] Kleine-Budde K, Müller R, Kawohl W, Bramesfeld A, Moock J und Rössler W. „The cost of depression - a cost analysis from a large database“. In: *Journal of Affective Disorders* 147.1 (Mai 2013), S. 137–43.
- [8] Wagner CJ, Metzger FG, Sievers C, Marschall U, L’hoest H, Stollenwerk B und Stock S. „Depression-related treatment and costs in Germany: do they change with comorbidity? A claims data analysis“. In: *Journal of Affective Disorders* 193 (2016), S. 257–266.

- [9] Zaprutko T, Göder R, Kus K, Pałys W, Rybakowski F und Nowakowska E. „The economic burden of inpatient care of depression in Poznan (Poland) and Kiel (Germany) in 2016“. In: *PloS One* 13.6 (2018), e0198890.
- [10] Schneider F, Härter M, Brand S, Sitta P, Menke R, Hammer-Filipiak U, Kudling R, Heindl A, Herold K, Frommberger U et al. „Adherence to guidelines for treatment of depression in in-patients“. In: *The British Journal of Psychiatry* 187.5 (2005), S. 462–469.
- [11] Chen P, Furukawa TA, Shinohara K, Honyashiki M, Imai H, Ichikawa K, Caldwell DM, Hunot V und Churchill R. „Quantity and quality of psychotherapy trials for depression in the past five decades“. In: *Journal of Affective Disorders* 165 (2014), S. 190–195.
- [12] Mohr DC, Spring B, Freedland KE, Beckner V, Arean P, Hollon SD, Ockene J und Kaplan R. „The selection and design of control conditions for randomized controlled trials of psychological interventions“. In: *Psychotherapy and Psychosomatics* 78.5 (2009), S. 275–84.
- [13] Cuijpers P, Straten A van, Bohlmeijer E, Hollon S und Andersson G. „The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size“. In: *Psychological Medicine* 40.2 (Feb. 2010), S. 211–23.
- [14] Mohr DC, Ho J, Hart TL, Baron KG, Berendsen M, Beckner V, Cai X, Cuijpers P, Spring B, Kinsinger SW et al. „Control condition design and implementation features in controlled trials: a meta-analysis of trials evaluating psychotherapy for depression“. In: *Translational Behavioral Medicine* 4.4 (2014), S. 407–423.
- [15] Cuijpers P, Cristea IA, Karyotaki E, Reijnders M und Huibers MJ. „How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence“. In: *World Psychiatry* 15.3 (Okt. 2016), S. 245–258.
- [16] Cuijpers P, Karyotaki E, Reijnders M und Ebert D. „Was Eysenck right after all? A reassessment of the effects of psychotherapy for adult depression“. In: *Epidemiology and Psychiatric Sciences* 28.1 (2019), S. 21–30.
- [17] Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A und Dobson KS. „A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments“. In: *The Canadian Journal of Psychiatry* 58.7 (2013), S. 376–385.
- [18] Furukawa TA, Weitz ES, Tanaka S, Hollon SD, Hofmann SG, Andersson G, Twisk J, DeRubeis RJ, Dimidjian S, Hegerl U et al. „Initial severity of depression and efficacy of cognitive-behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials“. In: *The British Journal of Psychiatry* 210.3 (2017), S. 190–196.

- [19] Barth J, Munder T, Gerger H, Nüesch E, Trelle S, Znoj H, Jüni P und Cuijpers P. „Comparative Efficacy of Seven Psychotherapeutic Interventions for Patients with Depression: A Network Meta-Analysis“. In: *PLoS Medicine* 10.5 (2013), e1001454.
- [20] Cuijpers P, Clignet F, Meijel B van, Straten A van, Li J und Andersson G. „Psychological treatment of depression in inpatients: a systematic review and meta-analysis“. In: *Clinical Psychology Review* 31.3 (2011), S. 353–360.
- [21] Borenstein M, Hedges LV, Higgins JP und Rothstein HR. *Introduction to meta-analysis*. Chichester: John Wiley & Sons, 2011.
- [22] Schefft C, Guhn A, Brakemeier EL, Sterzer P und Köhler S. „Efficacy of inpatient psychotherapy for major depressive disorder: a meta-analysis of controlled trials“. In: *Acta Psychiatrica Scandinavica* 139.4 (2018), S. 322–335.
- [23] Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savović J, Schulz KF, Weeks L und Sterne JA. „The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials“. In: *British Medical Journal* 343 (2011), S. d5928.
- [24] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2013. URL: <http://www.R-project.org/> (besucht am 30.03.2019).
- [25] Schwarzer G. „meta: An R package for meta-analysis“. In: *R News* 7 (2007), S. 40–45. URL: http://cran.r-project.org/doc/Rnews/Rnews_2007-3.pdf (besucht am 30.03.2019).
- [26] Schwarzer G, Carpenter JR und Rücker G. *Meta-Analysis with R*. Use R! Cham: Springer International Publishing, 2015.
- [27] Hedges LV. „Distribution theory for Glass’s estimator of effect size and related estimators“. In: *Journal of Educational Statistics* 6.2 (1981), S. 107–128.
- [28] Cohen J. *Statistical power analysis for the behavioral sciences*. 2. Aufl. Hillsdale, NJ: Erlsdale, 1988.
- [29] DerSimonian R und Laird N. „Meta-analysis in clinical trials“. In: *Controlled Clinical Trials* 7.3 (1986), S. 177–188.
- [30] Higgins JP, Thompson SG, Deeks JJ und Altman DG. „Measuring inconsistency in meta-analyses“. In: *British Medical Journal* 327.7414 (2003), S. 557.
- [31] Higgins JP und Thompson SG. „Controlling the risk of spurious findings from meta-regression“. In: *Statistics in medicine* 23.11 (2004), S. 1663–1682.
- [32] Murad MH, Montori VM, Ioannidis JP, Jaeschke R, Devereaux P, Prasad K, Neumann I, Carrasco-Labra A, Agoritsas T, Hatala R et al. „How to read a systematic review and meta-analysis and apply the results to patient care: users’ guides to the medical literature“. In: *JAMA* 312.2 (2014), S. 171–179.

- [33] Begg CB und Mazumdar M. „Operating characteristics of a rank correlation test for publication bias“. In: *Biometrics* 50.4 (1994), S. 1088–101.
- [34] Higgins JP und Green S, Hrsg. *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0*. The Cochrane Collaboration, 2011. URL: [www . handbook . cochrane . org](http://www.handbook.cochrane.org) (besucht am 30.03.2019).
- [35] Furukawa TA und Leucht S. „How to obtain NNT from Cohen’s d: comparison of two methods“. In: *PloS One* 6.4 (2011), e19070.
- [36] Button KS, Kounali DZ, Thomas L, Wiles NJ, Peters TJ, Welton NJ, Ades AE und Lewis G. „Minimal clinically important difference on the Beck Depression Inventory-II according to the patient’s perspective“. In: *Psychological medicine* 45.15 (2015), S. 3269–3279.
- [37] Klein JP, Erkens N, Schweiger U, Kriston L, Bausch P, Zobel I, Hautzinger M, Schoepf D, Serbanescu I, Bailer J et al. „Does childhood maltreatment moderate the effect of the cognitive behavioral analysis system of psychotherapy versus supportive psychotherapy in persistent depressive disorder?“ In: *Psychotherapy and Psychosomatics* 87.1 (2018), S. 46–48.
- [38] Correll CU, Rubio JM, Inczedy-Farkas G, Birnbaum ML, Kane JM und Leucht S. „Efficacy of 42 pharmacologic cotreatment strategies added to antipsychotic monotherapy in schizophrenia: systematic overview and quality appraisal of the meta-analytic evidence“. In: *JAMA Psychiatry* 74.7 (2017), S. 675–684.
- [39] Kraemer HC und Kupfer DJ. „Size of treatment effects and their importance to clinical research and practice“. In: *Biological psychiatry* 59.11 (2006), S. 990–996.

Eidesstattliche Erklärung

Ich, Cora SCHEFFT, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Wirksamkeit stationärer Psychotherapie bei Depressionen - eine Metaanalyse kontrollierter Studien“ selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe. Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen werden von mir verantwortet.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem Betreuer, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors www.icmje.org) zur Autorenschaft eingehalten. Ich erkläre ferner, dass mir die Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis bekannt ist und ich mich zur Einhaltung dieser Satzung verpflichte.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§156,161 des Strafgesetzbuches) sind mir bekannt und bewusst.

Datum:

Unterschrift:

Ausführliche Anteilserklärung an der erfolgten Publikation

Schefft C, Guhn A, Brakemeier EL, Sterzer P, und Köhler S. „Efficacy of inpatient psychotherapy for major depressive disorder: a meta-analysis of controlled trials“. In: Acta Psychiatrica Scandinavica 139.4 (2018), S. 322–335.

Beitrag im Einzelnen: Die Doktorandin Cora Schefft hat die hier vorliegende Publikation selbstständig erstellt. Sie hat nach Rücksprache mit ihrem Betreuer die Fragestellung eigenständig gesucht, das Analysekonzept der Arbeit entworfen sowie die Methodik ausgearbeitet. Die sehr umfangreiche Literaturrecherche sowie die ausführliche, detaillierte Neubewertung der eingeschlossenen Studien wurden von ihr selbstständig durchgeführt und die für die Metaanalyse verwendbaren Daten durch sie ausgewählt. Um die Durchführung der Metaanalyse und Neubewertung aller verwendeten Artikel zu gewährleisten, hat sie in Korrespondenz mit den Autoren der analysierten Artikel zusätzliche, von ihnen nicht veröffentlichte Daten erhoben und in die eigene Auswertung eingebaut. Frau Schefft hat sich die statistischen Analysen selbstständig angeeignet, durchgeführt und mit Unterstützung des Betreuers interpretiert. Die komplexe statistische Auswertung der Daten zur Metaanalyse wurde nach eigenständiger Erstellung eines Analyseplans durchgeführt. Das Manuskript wurde von ihr komplett eigenständig entworfen und verfasst. Herr PD Dr. Stephan Köhler und Herr Prof. Dr. Philipp Sterzer sowie Dr. Anne Guhn haben die Arbeit und die Erstellung des Manuskripts begleitet.

Unterschrift, Datum und Stempel des betreuenden Hochschullehrers

Unterschrift der Doktorandin

Auszug aus der Journal Summary List

Journal Data Filtered By: **Selected JCR Year: 2017** Selected Editions: SCIE,
Selected Categories: **"PSYCHIATRY"** Selected Category
Scheme: WoS

Gesamtanzahl: 142 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	World Psychiatry	4,055	30.000	0.010540
2	JAMA Psychiatry	8,414	16.642	0.044550
3	Lancet Psychiatry	3,223	15.233	0.015210
4	AMERICAN JOURNAL OF PSYCHIATRY	42,369	13.391	0.037870
5	PSYCHOTHERAPY AND PSYCHOSOMATICS	3,597	13.122	0.005520
6	BIOLOGICAL PSYCHIATRY	42,494	11.982	0.056910
7	MOLECULAR PSYCHIATRY	18,460	11.640	0.047200
8	JOURNAL OF NEUROLOGY NEUROSURGERY AND PSYCHIATRY	29,695	7.144	0.032980
9	SCHIZOPHRENIA BULLETIN	15,697	6.944	0.027700
10	NEUROPSYCHOPHARMACOLOGY	24,537	6.544	0.042870
11	JOURNAL OF CHILD PSYCHOLOGY AND PSYCHIATRY	18,604	6.486	0.023410
12	JOURNAL OF THE AMERICAN ACADEMY OF CHILD AND ADOLESCENT PSYCHIATRY	19,482	6.250	0.019260
13	ADDICTION	18,607	5.953	0.028990
14	BRITISH JOURNAL OF PSYCHIATRY	24,481	5.867	0.022960
15	Epidemiology and Psychiatric Sciences	950	5.684	0.003550
16	PSYCHOLOGICAL MEDICINE	23,080	5.475	0.039400
17	JOURNAL OF PSYCHIATRY & NEUROSCIENCE	2,989	5.182	0.004700
18	AUSTRALIAN AND NEW ZEALAND JOURNAL OF PSYCHIATRY	6,624	5.084	0.008440
19	DEPRESSION AND ANXIETY	7,923	5.043	0.015870
20	ACTA PSYCHIATRICA SCANDINAVICA	12,498	4.984	0.010890
21	JOURNAL OF PSYCHOPHARMACOLOGY	5,808	4.738	0.010900
22	PSYCHONEUROENDOCRINOLOGY	16,507	4.731	0.030420
23	Translational Psychiatry	5,384	4.691	0.021220
24	BIPOLAR DISORDERS	5,070	4.490	0.007870
25	CURRENT OPINION IN PSYCHIATRY	3,675	4.266	0.006830
26	JOURNAL OF CLINICAL PSYCHIATRY	18,677	4.247	0.020820
27	CNS DRUGS	4,364	4.206	0.007540
28	PROGRESS IN NEURO-PSYCHOPHARMACOLOGY & BIOLOGICAL PSYCHIATRY	9,823	4.185	0.013170
29	EUROPEAN NEUROPSYCHOPHARMACOLOGY	6,920	4.129	0.015110
29	EUROPEAN PSYCHIATRY	4,876	4.129	0.007890
31	JOURNAL OF PSYCHIATRIC RESEARCH	14,397	4.000	0.022480

Anhang A

Originalpublikation und Supporting Information

- A.1** Schefft C, Guhn A, Brakemeier EL, Sterzer P, und Köhler S. „Efficacy of inpatient psychotherapy for major depressive disorder: a meta-analysis of controlled trials“. In: *Acta Psychiatrica Scandinavica* 139.4 (2018), S. 322–335. <https://doi.org/10.1111/acps.12995>

Anhang B

Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Anhang C

Publikationsliste

- 1 Schefft C, Guhn A, Brakemeier EL, Sterzer P und Köhler S. „Efficacy of inpatient psychotherapy for major depressive disorder: a meta-analysis of controlled trials“. In: *Acta Psychiatrica Scandinavica* 139.4 (2018), S. 322–335.
Impact Factor: 4.984
- 2 Schefft C, Kilarski LL, Bschor T und Koehler S. „Efficacy of adding nutritional supplements in unipolar depression: A systematic review and meta-analysis“. In: *European Neuropsychopharmacology* 27.11 (2017), S. 1090–1109.
Impact Factor: 4.129

Danksagung

Ich danke meinen Betreuern PD Dr. Stephan Köhler und Prof. Dr. Philipp Sterzer, sowie Dr. Anne Guhn und Prof. Dr. Eva-Lotta Brakemeier. Meiner Familie, Eric Schulz und meinen Freunden.