Untersuchung eines Publikationsbias in Meta-Analysen zur Wirksamkeit psychotherapeutischer Interventionen bei Essstörungen, Depression und Schizophrenie

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Helen Niemeyer
geboren in Minden

Düsseldorf, Oktober 2011

Aus dem Institut für Experimentelle Psychologie
Der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit Genehmigung
Der Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent:	Prof. Dr. Reinhard Pietrowsky
Koreferent:	Prof. Dr. Jochen Musch

Tag der mündlichen Prüfung: 18.11.2011

Danksagung

Für die wertvolle Unterstützung beim Schreiben dieser Doktorarbeit möchte ich mich ganz herzlich bei meinem Doktorvater Professor Reinhard Pietrowsky bedanken, der viel Zeit und Rat investiert hat, und ebenfalls bei meinem Zweitgutachter Professor Jochen Musch, der intensiv am Verfassen der Artikel mitgearbeitet hat. Weiterhin gilt mein herzlicher Dank allen meinen Kollegen aus der Abteilung für Klinische Psychologie, die mich in den vier Jahren begleitet haben und mir immer mit Rat und Tat zur Seite standen, insbesondere Frank, Melanie, Friederike und Anja. Natürlich gilt mein großer Dank auch unseren studentischen Hilfskräften, Theresa, Annett und Doris, die mir sehr viel Arbeit abgenommen haben.

Zudem möchte ich mich sehr herzlich bei meinen Eltern bedanken, die mich immer unterstützt haben, und leider aufgrund des hohen Arbeitspensums häufig auf meine Anwesenheit verzichten mußten. Ich möchte mich auch ganz besonders bei meinen Freunden Iris, Anne und Sabrina bedanken, die mir beim Korrekturlesen geholfen haben und meine Sorgen rund um die Doktorarbeit mit mir besprochen und geteilt haben.

Inhalt

Zusammenfassung

Das Thema der vorliegenden Arbeit ist die umfassende statistische Prüfung eines Publikationsbias in Meta-Analysen zur Wirksamkeit psychotherapeutischer Interventionen bei den Störungsbildern Bulimia Nervosa, Anorexia Nervosa, Binge Eating Disorder, Depression und Schizophrenie, sowie in Meta-Analysen zur Wirksamkeit von Präventionsprogrammen bei Essstörungen und Depressionen. Ein Publikationsbias ist charakterisiert als Veröffentlichung von signifikanten Studienergebnissen, im Gegensatz zu nicht-signifikanten Ergebnissen. Infolgedessen kann der Berechnung von Effektstärken in Meta-Analysen eine verzerrte Stichprobe von Primärstudien zugrunde liegen, wodurch die Wirksamkeit therapeutischer Interventionen überschätzt würde. Ein Publikationsbias stellt eine der größten Bedrohungen der Validität von Meta-Analysen dar. Empirische Belege für die Existenz eines Publikationsbias liegen aus anderen Forschungsbereichen vor, in Meta-Analysen der Wirksamkeitsforschung psychotherapeutischer Interventionen ist eine umfassende Prüfung des Vorhandenseins eines Publikationsbias und seiner Auswirkungen auf die Höhe der mittleren Effektstärken jedoch nicht durchgeführt worden. In der vorliegenden Arbeit wurden zu den genannten Störungsbereichen sämtliche Meta-Analysen, deren statistische Daten einer Überprüfung zugänglich waren und in denen homogene Effektstärken berichtet wurden, mittels Begg und Mazumdars Rangkorrelationsmethode, Eggers Regressionsanalyse und der Trim und Fill Methode auf einen Publikationsbias untersucht. Es zeigte sich, dass ein Publikationsbias nur in geringem Ausmaß vorlag, und die Wirksamkeit der untersuchten Interventionen kaum überschätzt wurde. Für nur vier der untersuchten Interventionen änderte sich die Wirksamkeitseinschätzung nach der Korrektur um den Einfluss eines Publikationsbias. Zuvor bei Schizophrenie nicht wirksame Familieninterventionen wurden nach der Korrektur als wirksam eingeschätzt, Psychotherapie zeigte sich bei depressiven Patienten mit und ohne komorbide Persönlichkeitsstörung als gleich wirksam, Psychopharmaka als ebenso wirksam bei Depression wie Psychotherapie, und kognitive Verhaltenstherapie anderen Therapien bei Essstörungen überlegen. Mit Ausnahme einiger Interventionen für Schizophrenie und Essstörungen, die bereits vor der Korrektur um den Einfluss eines Publikationsbias nicht-signifikant ausfielen, ist die Wirksamkeit der Therapieverfahren substantiell gegeben. Die vorliegende Arbeit trägt zu einer Erhöhung der Validität und Robustheit der empirischen Evidenzen bei. Als wichtigste Konsequenz für die Anwendung der geprüften Interventionen in der klinischen Praxis ist eine erhöhte Sicherheit und Verlässlichkeit in deren Wirksamkeit aus den vorliegenden Ergebnissen abzuleiten.

Abstract

The topic of the present study is the comprehensive statistical test of publication bias in meta-analyses of the effectiveness of psychotherapeutic interventions for bulimia nervosa, anorexia nervosa, binge eating disorder, depression and schizophrenia, as well as meta-analyses of the effectiveness of prevention programs for eating disorders and depression. Publication bias is characterized as the preferential publication of significant rather than non-significant results. Due to this practice the calculation of effect sizes in meta-analyses may be based on a biased sample of primary studies, which either does not include null or negative results at all or under represents such results, leading to an overestimation of the effectiveness of therapeutic interventions. Publication bias thus poses one of the greatest threats to the validity of meta-analyses. While empirical evidence for the existence of publication bias is available in other areas of research, no comprehensive examination of its existence and its effects on the average effect size in meta-analyses of research on the effectiveness of psychotherapeutic interventions has been conducted systematically. Therefore, all meta-analyses of the three disorders for which sufficient statistical details were available and that reported homogeneous effect sizes were reassessed. Begg and Mazumdar´s rank correlation method, Egger´s regression analysis and the Trim and Fill procedure were applied. We found only a small degree of publication bias in therapy research, and the effectiveness of the interventions studied was rarely overestimated. The evaluation of efficacy changed for only four interventions after being corrected for publication bias. Family interventions for schizophrenia were assessed as being efficacious after correcting for publication bias, psychotherapy was shown to be of equal efficacy for depressed patients with and without comorbid personality disorder, the correction led to an equal efficacy of psychotherapy and pharmacotherapy for depression, and cognitive behavioral therapy was superior to other psychotherapies for eating disorders. In sum, the effectiveness of the examined interventions can be viewed as substantial, except for a few interventions for schizophrenia and eating disorders, which had been non-significant before the correction, as well. The present study bolsters the validity and robustness of the empirical evidence for the effectiveness of therapeutic interventions. The most important conclusion to be drawn from the present results is the increase in safety and reliability of the application of the examined interventions in clinical practice.

Einführung

Zur Überprüfung der Wirksamkeit psychotherapeutischer Intervention hat das Konzept der evidenzbasierten Psychotherapieforschung zunehmend an Relevanz gewonnen (Fydrich & Schneider, 2007). In der evidenzbasierten Wirksamkeitsforschung basieren Effektivitätsnachweise für therapeutische Interventionen bei psychischen Störungen zum einen auf den empirischen Ergebnissen einzelner Studien. Zum anderen ist eine systematische Zusammenfassung und statistische Integration der einzelnen Studienergebnisse notwendig, da mittlerweile zu vielen Störungsbildern und Therapieverfahren sehr große Datenmengen vorliegen. Zur Menge der Daten kommt hinzu, dass die Ergebnisse einzelner Studien oftmals heterogen sind. Deshalb kann ein valider Überblick über sämtliche Studienergebnisse von einzelnen Forschern ohne quantitative Integrationsmethoden nicht mehr objektiv geleistet werden (Hopewell, Clarke & Mallett, 2005; Rustenbach, 2003). Die unter dem Sammelbegriff der Meta-Analyse zusammengefassten Methoden der Studienintegration und Effektstärkenberechnung sind deshalb notwendig, um Einzelergebnisse in einem Forschungsbereich zu aussagekräftigen Gesamtwerten zu aggregieren (Borenstein, Hedges, Higgins & Rothstein, 2009; Rustenbach, 2003). In Meta-Analysen werden Effektstärken zur Bestimmung der Wirksamkeit von therapeutischen Interventionen berechnet, die folglich auf größeren Stichproben basieren und idealerweise den vollständigen Wissensstand in einem Forschungsfeld zu einer Hypothese zusammenfassen. In der evidenzbasierten Forschung wird somit im ersten Schritt die Wirksamkeit der Interventionen in einzelnen empirischen Studien untersucht. Diese Studien werden in einem zweiten Schritt zur Erhöhung der Teststärke und Validität in Meta-Analysen integriert. Die Güte empirischer Evidenz wird hierarchisch geordnet, wobei auf kontrollierten, randomisierten Studien basierende Meta-Analysen die höchste Evidenzklasse darstellen: Meta-analytische Effektstärken, als kumulierte Evidenz, werden als höchste und beste Ebene wissenschaftlicher Evidenz betrachtet (Fydrich & Schneider, 2007; Slade & Priebe, 2001).

Meta-analytische Methoden sind durch eine hohe Durchführungsobjektivität, Ergebnisreliabilität und Ergebnisvalidität gekennzeichnet (Rustenbach, 2003). Diese hängen allerdings von der korrekten Anwendung der Methoden ab. Der Validität meta-analytischer Ergebnisse liegt dabei insbesondere die Vollständigkeit der eingeschlossenen Daten zugrunde. Eine Meta-Analyse ist als Beobachtungsstudie aufzufassen, da keine zufällige Ziehung der einzuschließenden Primärstudien, analog zur Stichprobenbildung bei einer randomisierten Studie, erfolgen kann (Hunter & Schmidt, 2000; Wolf, 1990). Aus diesem Grund müssen

sämtliche Studienergebnisse, sowohl publizierte als auch unpublizierte, berücksichtigt werden, um die Repräsentativität meta-analytischer Ergebnisse zu gewährleisten. Die größte Verzerrungsquelle für die Validität wird von vielen Forschern folglich im Publikationsbias gesehen (Rothstein, Sutton & Borenstein, 2005; Rustenbach, 2003). Der Publikationsbias ist ein systematischer Bias, der durch die vorwiegende Veröffentlichung von Studien mit signifikanten Ergebnissen, im Gegensatz zu Studien mit nicht-signifikanten und Null-Ergebnissen, charakterisiert ist (Hopewell et al., 2005; Rothstein et al., 2005). Der Begriff Publikationsbias bezeichnet somit die selektive Verfügbarkeit von Studien basierend auf der Größe und Richtung ihrer Effekte. Aufgrund der Abhängigkeit der Veröffentlichung von der Richtung und Signifikanz der Ergebnisse ist sie ein nicht-zufälliges Ereignis. Als eine weitere Ursache des Publikationsbias wird die vorrangige Veröffentlichung von „positiven" Ergebnissen diskutiert. Positive Ergebnisse umfassen solche, die der Hypothese des Forschers bei der Durchführung der Studie entsprechen oder die eher unerwartete und somit besonders berichtenswerte Ergebnisse darstellen (Dickersin, 2005). Werden für eine Veröffentlichung die statistische Signifikanz und die Größe der Effekte sowie ihre Hypothesenkonformität als ausschlaggebend betrachtet, und in Folge lediglich publizierte Forschungsergebnisse anderen Wissenschaftlern zugängig gemacht, führt dieser systematische Bias zu einer Überschätzung der Wirksamkeit der untersuchten Intervention. Da bei der Studienaggregation in Meta-Analysen aufgrund der leichteren Zugänglichkeit hauptsächlich publizierte Studien eingeschlossen werden, liefert eine Meta-Analyse keine umfassende und korrekte Wiedergabe der Studienlage bzw. Wissenslage mehr. Diese Verzerrung kann in einer Überhöhung der mittleren Effektstärke resultieren (Dickersin, 2005; Rustenbach, 2003). Die Validität der Ergebnisse einer Meta-Analyse erfordert daher die Identifizierung einer unverzerrten und vollständigen Studienstichprobe, welche bei Vorliegen eines Publikationsbias nicht mehr gegeben ist. Die Grundlage der Durchführung einer Meta-Analyse ist deshalb die Identifizierung und statistische Zusammenfassung der relevanten Primärstudien in einer Weise, in der ein Publikationsbias ausgeschlossen wird (Hopewell et al., 2005; Mulrow, 1994).

Empirische Belege zum Publikationsbias liegen direkt und indirekt vor (Sutton, 2005). Unter die Kategorie der direkten Belege fallen Befragungen von Forschern zu den Gründen, ihre Studien nicht zur Publikation einzureichen. Auch die Überprüfung der Publikation der Ergebnisse von Studien, die zu Beginn der Studiendurchführung registriert wurden, die zeigt, dass nicht-signifikante Ergebnisse seltener veröffentlicht werden, gehört zu den direkten Belegen. Die Nichtveröffentlichung nicht-signifikanter Ergebnisse geht sowohl auf die Praxis

der Forscher als auch der Herausgeber zurück (Dickersin, 2005). Forscher reichen negative Ergebnisse seltener ein, bereits in Erwartung einer Ablehnung. Gleichsam lehnen Herausgeber die Veröffentlichung von Null- oder Negativeffekten als uninteressant oder nicht informativ ab. Zu der indirekten empirischen Evidenz eines Publikationsbias zählen die Quantifizierung der Überrepräsentation signifikanter Ergebnisse in der veröffentlichten Literatur, negative Korrelationen zwischen Studiengröße und Behandlungseffekten in veröffentlichten Studien sowie die statistische Überprüfung in Meta-Analysen (Sutton, 2005).

Indirekte Belege für eine selektive Veröffentlichungspraxis liegen seit über 50 Jahren vor. Sterling führte bereits 1959 eine Untersuchung publizierter Studienergebnisse durch. Er konnte zeigen, dass der Großteil der in der Psychologie veröffentlichten Studien signifikante Ergebnisse aufwies (Sterling, 1959). Spätere Studien in den Sozial- und Verhaltenswissenschaften (Smith, 1980a, 1980b; White, 1982) sowie in der Medizin (Ernst & Pittler, 1997; Moscati, Jehle, Ellis, Fiorello & Landi, 1994; Vickers, Gloyal, Harland & Rees, 1998) konnten dieses Ergebnis in ihren Forschungsfeldern replizieren. Weiterhin konnte in mehreren Studien gezeigt werden, dass Studien, in denen die Nullhypothese zurückgewiesen wurde, häufiger zur Veröffentlichung eingereicht wurden (Coursol & Wagner, 1986; Greenwald, 1975; Shadish, Doherty & Montgomery, 1989). Direkte Belege liegen aus zwei weiteren Studien vor (Epstein, 1990; Mahoney, 1977), in denen Manuskripte zur Veröffentlichung bei Fachzeitschriften eingereicht wurden, die sich theoretisch und methodisch stark ähnelten. Allerdings wurde die Hypothesenkonformität der Ergebnisse von den Autoren gezielt geändert, so dass jeweils Manuskripte mit positiven sowie mit negativen Ergebnissen bei Fachzeitschriften eingereicht wurden. Beide Studien ergaben, dass diejenigen Manuskripte mit positiven Ergebnissen, definiert als hypothesenkonform, häufiger zur Veröffentlichung angenommen wurden. Dickersin, Chan, Chalmers, Sacks und Smith (1987) fanden zudem in einer Befragung von Forschern heraus, dass der häufigste Grund, Studienergebnisse nicht zur Veröffentlichung einzureichen, das Vorliegen negativer und nicht-signifikanter Ergebnisse war. In einem späteren Review zur Überprüfung der Veröffentlichung der Studienergebnisse von sechs Kohortenstudien aus dem Feld der Biowissenschaften konnte Dickersin das Ergebnis replizieren, dass der Hauptgrund für Nichtveröffentlichungen in dem Vorliegen von negativen und Nullergebnissen lag (Dickersin, 1997). Scherer, Langenberg und van Elm (2005) fanden in einer neueren Untersuchung von Studien der Biomedizin ebenfalls, dass signifikante Ergebnisse mit größerer Wahrscheinlichkeit veröffentlicht werden. Cooper, DeNeve und Charlton (1997) fanden heraus, dass von 117 psychologischen Studien in den Jahren 1986 bis 1988 62% statistisch

signifikante Ergebnisse aufwiesen. Während 74% dieser Studien mit signifikanten Ergebnissen zur Veröffentlichung eingereicht wurden, lag der Prozentanteil bei den Studien mit nicht-signifikanten Ergebnissen bei nur 4%. Glass, McGaw und Smith (1981) kamen nach einer Untersuchung von neun zwischen 1976 und 1980 veröffentlichten Meta-Analysen aus dem Bereich der Psychologie zu dem Ergebnis, dass die mittlere Effektstärke eingeschlossener publizierter Studien mit einem Cohens *d* von 0.64 um 33% größer war als die der unpublizierten Studien mit $d = 0.48$. Lipsey und Wilson (1993) führten eine ähnliche Untersuchung an einer größeren Stichprobe von Meta-Analysen aus dem Bereich der Therapieforschung durch. Sie untersuchten 92 Meta-Analysen und fanden ebenfalls, dass die Effektstärken publizierter Studien im Mittel um 0.14 Standardabweichungen größer waren als die unpublizierter Studien. Sterne et al. (2002) konnten ebenfalls bestätigen, dass in 122 Meta-Analysen der Cochrane Database of Systematic Reviews die mittleren Effektstärken der publizierten Studien signifikant größer waren als die der unpublizierten. Zusammenfassend belegen diese Untersuchungen, dass systematische Unterschiede zwischen den Ergebnissen veröffentlichter und unveröffentlichter Studien vorliegen.

Die Ergebnisse der dargestellten indirekten und direkten Untersuchungen weisen zusammengenommen darauf hin, dass ein Publikationsbias in einem Großteil der untersuchten Meta-Analysen vorliegt. Sie unterstreichen die Notwendigkeit, unveröffentlichte Studien in Meta-Analysen einzuschließen (Hopewell et al., 2005). Da unveröffentlichte Studien jedoch schwer zugänglich sind, und somit ein hoher Zeitaufwand für die diesbezügliche Recherche, bei gleichzeitig geringen Erfolgsaussichten, notwendig ist, werden in der Mehrzahl der Meta-Analysen hauptsächlich veröffentlichte Studien eingeschlossen. Außerdem wird der Effekt eines möglichen Publikationsbias auf die Ergebnisse in den meisten Meta-Analysen nicht berücksichtigt. Sensitivitätsanalysen zur Überprüfung des Vorhandseins und des Einflusses eines Publikationsbias auf die mittlere Effektstärke sollten jedoch als methodischer Standard in jeder Meta-Analyse durchgeführt werden.

Statistische Methoden zur Überprüfung eines Publikationsbias sind erst nach der Verbreitung der quantitativen Techniken der Meta-Analyse in den letzten Jahrzehnten entwickelt geworden, da es einhergehend mit der Verwendung quantitativer meta-analytischer Methoden erst möglich wurde, den Einfluss eines Publikationsbias auch empirisch nachzuweisen und seine Auswirkungen auf die mittlere Effektstärke quantitativ zu erfassen. Der Fokus auf die objektiven quantitativen Ergebnisse von Meta-Analysen hat somit dazu

geführt, dass dem Problem der Publikationsverzerrung statistisch begegnet werden konnte (Rothstein et al., 2005).

Die statistischen Verfahren zur Überprüfung eines Publikationsbias, die auch in dieser Untersuchung angewendeten wurden, basieren auf dem Funnel Plot, dem wohl bekanntesten Verfahren (Light & Pillemer, 1984). Der Funnel Plot ist ein Streudiagramm, in dem die Effektstärken der einzelnen Primärstudien einer Meta-Analyse auf der Abszisse und ein Maß ihrer Präzision auf der Ordinate abgetragen werden. Große Studien verteilen sich dabei nahe der mittleren Effektstärke oben im Graphen, da sie präziser sind, während die kleineren Studien aufgrund ihrer geringeren Präzision am Boden des Graphen weiter streuen. Die Verteilung verengt sich dadurch zur Spitze des Graphen hin. In Abwesenheit eines Publikationsbias ist der Funnel Plot symmetrisch, da sich die Primärstudieneffekte auf jedem Präzisionslevel symmetrisch um den mittleren Effektstärkeschätzer verteilen. Dadurch sieht der Funnel Plot wie ein umgekehrter Trichter aus, woher er seinen Namen bezieht. Asymmetrie in der Verteilung der Studien ist ein Indikator für einen Publikationsbias. Kleinere Studien erfüllen das Kriterium der statistischen Signifikanz leichter und werden mit höherer Wahrscheinlichkeit publiziert, wenn sie große Effektstärken aufweisen. Fehlen unpublizierte Studien mit null- oder negativen Effekten in der Meta-Analyse, weil diese nicht publiziert wurden, entsteht an der Stelle der Verteilung eine Lücke, an der sich diese, meist kleineren Studien befinden sollten. Dadurch ist die Konzentration von Studien auf der rechten Seite der mittleren Effektstärke im Graphen höher, auf der die größeren Effektstärken abgebildet sind, und eine Asymmetrie ist gegeben (Sterne, Becker & Egger, 2005). Der Hauptnachteil dieser Methode ist die Subjektivität ihrer Interpretation, da eine Asymmetrie der Verteilung nur visuell eingeschätzt werden kann und somit unreliabel ist. Erschwerend kommt hinzu, dass die verschiedenen Präzisionsmaße, die verwendet werden können, zu sehr unterschiedlichen Verteilungsformen führen. Um dieses Problem zu lösen, sind quantitative und somit objektivere Methoden zur Überprüfung eines Publikationsbias entwickelt worden, die auf dem Funnel Plot basieren und im Folgenden dargestellt werden (Duval, 2005; Sterne & Egger, 2005). Die Verwendung dieser statistischen Methoden zur Überprüfung eines möglichen Publikationsbias ist aufgrund der Objektivität und Replizierbarkeit immer der subjektiven visuellen Beurteilung eines Funnel Plots vorzuziehen.

Begg und Mazumdars Rangkorrelationstest (1994) ist ein non-parametrisches Verfahren zur Überprüfung eines Small-Study-Bias. Weisen kleinere Studien (die höhere Varianzen zeigen) größere Effektstärken auf, da kleine Studien mit großen Effekten eher

signifikant werden und deshalb mit größerer Wahrscheinlichkeit veröffentlicht werden, so wird sich eine positive Korrelation zwischen der Effektstärke und der Varianz zeigen. Dieser Sachverhalt wird als Small-Study-Bias bezeichnet. Im Rangkorrelationstest werden den standardisierten Effektstärken und den Varianzen Ränge zugewiesen, und die Rangkorrelation (Kendall´s Tau) wird verwendet, um einen Zusammenhang zwischen beiden Maßen zu prüfen, indem die Anzahl konkordanter sowie diskordanter Ränge auf Signifikanz getestet wird. In Abwesenheit eines Small-Study- respektive Publikationsbias ist keine signifikante Korrelation zwischen den Rängen zu erwarten. Aus der statistischen Überprüfung lässt sich die Ursache der Korrelation natürlich nicht ableiten, aber ein Publikationsbias ist einer der wahrscheinlichsten Gründe für das Vorliegen eines solchen Small-Study-Bias (Sterne & Egger, 2005).

Ein weiteres Verfahren, Eggers lineare Regressionsanalyse (Egger, Davey Smith, Schneider & Minder, 1997), prüft die in einer Meta-Analyse eingeschlossenen Studien ebenfalls auf einen Small-Study-Bias. Dieses parametrische Verfahren weist eine höhere Teststärke als der non-parametrische Rangkorrelationstest auf. Die Standardnormalverteilung, die an ihrem Standardfehler standardisierte Effektstärke, wird auf ihre Präzision, die Inverse ihres Standardfehlers, zurückgeführt. Die Steigung der Regressionsgeraden spiegelt die Größe und Richtung des Effektes wieder, während der Punkt, an dem die Gerade die Ordinate schneidet, der so genannte Interzept, ein Maß der Asymmetrie darstellt: Verläuft die Regressionsgerade nicht durch den Ursprung des Koordinatensystems, zeigt dies eine Asymmetrie in der Verteilung an, wodurch der Interzept als ein Maß für die Stärke der Asymmetrie genutzt werden kann (Sterne & Egger, 2005).

Für beide Methoden, den Rangkorrelationstest sowie die Regressionsanalyse, wird empfohlen, zweiseitige Signifikanztests zu verwenden, da auch ein gegenteiliges Asymmetrie-Muster auftreten kann (Sterne & Egger, 2005). Andere Autoren argumentieren für eine einseitige Testung, die dem Sachverhalt angemessen ist, dass Studien mit nicht-signifikanten und negativen Effekten seltener veröffentlicht werden als solche mit signifikanten positiven Effekten, und ein Publikationsbias deshalb die mittlere Effektstärke eher verkleinern als vergrößern würde (Cuijpers, Smit, Bohlmeijer, Hollon & Andersson, 2010). Deshalb werden in der vorliegenden Arbeit sowohl ein- als auch zweiseitige Signifikanztests berichtet. Weiterhin sollten beide Methoden nur bei Datensätzen angewandt werden, die mindestens sechs Studien umfassen, da die Teststärke bei weniger Studien zu gering ausfällt (Egger et al., 1997).

Neben der reinen Prüfung des Vorliegens einer Asymmetrie adjustiert eine dritte Methode, Trim und Fill (Taylor & Tweedie, 2000a, 2000b), die mittlere Effektstärke bei Vorliegen eines Publikationsbias zusätzlich um dessen Einfluß und schätzt die Anzahl der fehlenden Studien. Eine korrigierte mittlere Effektstärke, bereinigt um den Einfluß des Publikationsbias, wird berechnet. Die Methode arbeitet nicht mit der Korrelation zwischen den Effektstärken und den Präzisionsmaßen, wie der Rangkorrelationstest und die Regressionsanalyse, sondern prüft, ob sich die einzelnen Primärstudieneffekte symmetrisch um die mittlere Effektstärke verteilen. Der mittlere Effektstärkeschätzer wird als Ausgangspunkt genommen, um den einzelnen Primärstudien Ränge zuzuweisen. Den Primärstudien werden positive Ränge zugeordnet, wenn sie größere Effektstärken als die mittlere Effektstärke aufweisen, und negative Ränge, wenn kleinere Effektstärken vorliegen. Gibt es weniger Studien mit kleinen Effektstärken, und weisen dadurch Studien mit großen Effektstärken keine ihnen entsprechenden „spiegelbildlichen“ Gegenstudien mit kleinen Effektstärken auf der anderen Seite des mittleren Effektes auf, liegt eine Ungleichverteilung, also eine Asymmetrie, vor. In einem ersten Schritt werden diese Studien mit großen Effektstärken aus der Verteilung entfernt. Zumeist sind dies die Studien mit den größten Effektstärken (Duval, 2005). In einem zweiten Schritt wird für die verbleibenden Studien eine neue mittlere Effektstärke berechnet, und anschließend werden die zuvor entfernten Studien wieder eingefüllt. Mittels eines iterativen Algorithmus werden diese Schritte solange wiederholt, bis eine Symmetrie der Verteilung erreicht ist. Meist verbleibt der Algorithmus nach wenigen Arbeitsschritten stabil. Sämtliche „getrimmten“ Studien sowie deren neu erstellte spiegelbildliche Gegenstudien mit negativen Effektstärken werden in die Verteilung „eingefüllt“, und das Programm gibt die Anzahl dieser fehlenden Studien an. Eine neue mittlere Effektstärke wird ausgegeben, die auf den beobachteten sowie den neu eingefüllten Studien basiert. Zumeist fällt diese Bias-korrigierte Effektstärke niedriger als die alte aus. Ein neues 95%-Konfidenzintervall wird ebenfalls berechnet. Die Reduktion der mittleren Effektstärke durch Trim und Fill kann auf Signifikanz getestet werden, indem geprüft wird, ob der ursprüngliche Effektschätzer noch innerhalb des neuen Konfidenzintervalls des um den Einfluss des Publikationsbias korrigierten Effektstärkeschätzers liegt. Zusätzlich muss überprüft werden, ob die Wirksamkeit der untersuchten Intervention nach der Korrektur gegeben ist. Dazu wird die neue Effektstärke auf Signifikanz, also auf ihre zufallskritische Differenz zu Null, getestet. Die Trim und Fill Methode sollte als Sensitivitätsanalyse betrachtet werden, in der die Robustheit der mittleren Effektstärke einer Meta-Analyse auf den Einfluss eines Publikationsbias hin überprüft wird. Die von dem Verfahren ausgegebene

Anzahl der fehlenden Studien sollte nicht als exakter Wert der Anzahl wirklich vorhandener unpublizierter Studien interpretiert werden (Duval, 2005).

Eine Asymmetrie in der Verteilung der Primärstudien-Effektstärken kann auch durch andere Ursachen als einen Publikationsbias hervorgerufen werden. Hier ist insbesondere eine Heterogenität der Primärstudien-Effektstärken zu nennen. Wenn nicht alle Primärstudien denselben zugrundeliegenden wahren Effekt messen, aber dennoch zu einem mittleren Effektstärkeschätzer zusammengefasst werden, entsteht Heterogenität in diesem Datensatz (Borenstein et al., 2009). Da Heterogenität ebenfalls zu einer Asymmetrie im Funnel Plot führt, ist die Anwendung der dargestellten, auf einer Homogenitätsannahme basierenden Methoden bei Heterogenität fehlerhaft (Ioannidis & Trikalinos, 2007; Sterne, Gavaghan & Egger, 2000). Die Homogenitätsvoraussetzung würde in diesem Fall verletzt, und eine Anwendung würde zu falsch positiven Ergebnissen führen (Terrin, Schmid, Lau & Olkin, 2003).

Weiterhin kann jeder Faktor, der sowohl mit der Studiengröße als auch der Höhe der Effektstärke zusammenhängt, zu einer Asymmetrie in der Verteilung führen (Sterne et al., 2005). Auch eine divergierende Qualität der eingeschlossenen Studien kann ein Grund für Asymmetrie sein. Letzten Endes kann eine Asymmetrie aber immer auch lediglich durch den Zufall bedingt sein. Aufgrund dessen sind die vorgestellten statistischen Methoden indirekte Werkzeuge, die keine endgültigen Rückschlüsse auf die Ursache der Asymmetrie zulassen (Egger et al., 1997). Dennoch ist ein Publikationsbias eine plausible Erklärung für ein Asymmetrie-Muster in einem homogenen Datensatz, in dem sich genau dort eine Lücke in der Verteilung der Primärstudien zeigt, wo sich Studien mit nicht-signifikanten Ergebnissen befindet würden.

Die erste umfassende statistische Überprüfung eines Publikationsbias in einer Stichprobe von Meta-Analysen führten Egger et al. (1997) mittels der Regressionsanalyse durch. Die Asymmetrie der Verteilung der Effektstärken der eingeschlossenen Primärstudien wurde in Meta-Analysen, die zwischen 1993 und 1996 in vier medizinischen Fachzeitschriften veröffentlicht wurden, sowie Meta-Analysen der Cochrane Database of Systematic Reviews statistisch geprüft. Einschlusskriterien waren ein dichotomes Effektstärke-Maß, sowie eine Mindestanzahl von sechs Primärstudien je Studienarm. 75 Meta-Analysen konnten eingeschlossen werden, und der Regressionstest wurde in 5 (13%) der Cochrane Meta-Analysen auf dem 10%-Niveau signifikant, sowie in 14 (38%) der Meta-Analysen, die den medizinischen Fachzeitschriften entnommen wurden.

In einer Erhebung von Sterne et al. (2000), in der neben der Regressionsanalyse (Egger et al., 1997) auch der Rangkorrelationstest von Begg und Mazumdar (1994) angewendet wurde, wurden Meta-Analysen aus acht medizinischen Fachzeitschriften untersucht, die zwischen 1993 und 1997 veröffentlicht wurden, und es galten dieselben Einschlusskriterien wie in der Analyse von Egger et al. (1997). Von den 87 geprüften Meta-Analysen ergab der Regressionstest (Egger et al., 1997) bei 21 (26.9%) ein signifikantes Ergebnis (10%- Signifikanzniveau), sowie der Rangkorrelationstest bei 10 (12.8%). Diese Ergebnisse weisen in dieselbe Richtung wie die der Studie von Egger et al. (1997).

Sutton, Duval, Tweedie, Abrams und Jones (2000) ergänzte die statistische Überprüfung von Meta-Analysen um die Anwendung der Trim und Fill Methode (Taylor & Tweedie, 2000a, 2000b). Achtundvierzig Meta-Analysen der Cochrane Database of Systematic Reviews, in denen ausschließlich randomisierte kontrollierte Studien eingeschlossen waren, und in die mindestens zehn Primärstudien eingingen, wurden getestet. Unter der Verwendung eines Fixed Effekt-Modells (FEM) zur Studienintegration fügte Trim und Fill bei 26 Meta-Analysen (54%) mindestens eine fehlende Studie ein. Wurde ein Random Effekt-Modell (REM) zugrunde gelegt, waren es 23 Meta-Analysen, in denen fehlende Studien eingefügt wurden (48%). Jedoch war die Reduktion der mittleren Effektstärke in fast allen Fällen nur minimal, und nur bei 3 (FEM) respektive 4 (REM) Meta-Analysen fiel die Reduktion der mittleren Effektstärke signifikant aus. Der Regressionstest (Egger et al., 1997) wurde bei 13 (27%) der Meta-Analysen auf dem 5%-Niveau signifikant, der Rangkorrelationstest bei 7 (15%) Meta-Analysen (Sutton, 2005).

Song, Khan, Dinnes und Sutton (2002) verwendeten ebenfalls Begg und Mazumdars Rangkorrelationstest (1994), Eggers Regressionsanalyse (Egger et al., 1997) und Taylor und Tweedies Trim und Fill Methode (2000a, 2000b) zur Überprüfung eines möglichen Publikationsbias in Meta-Analysen zu diagnostischen Tests aus dem Jahr 1999. Wie bei Egger et al. (1997) musste ein dichotomes Effektstärke-Maß berichtet werden, und es mussten mindestens sechs Primärstudien pro Arm eingeschlossen sein. Achtundzwanzig Meta-Analysen erfüllten die Einschlusskriterien. Der Rangkorrelationstest wurde bei 23 (82%) Meta-Analysen auf dem 5%-Niveau signifikant, der Regressionstest bei 12 (43%) Meta-Analysen, und Trim und Fill fügte unter Verwendung des REM bei 17 (61%) und unter Verwendung des FEM bei 21 Meta-Analysen (75%) fehlende Studien ein. Song et al. (2002) berechneten, bei wie vielen Meta-Analysen die mittlere Effektstärke um 11-20% überschätzt

worden war, und fanden dieses Ergebnis bei 7 (25%) der Meta-Analysen. Eine stärkere Überschätzung von 21-30% lag bei zwei (7%) Meta-Analysen vor.

Eine Überprüfung von 40 Meta-Analysen aus der Biologie (Jennions & Moller, 2002), in denen mindestens 8 Primärstudien pro Arm eingeschlossen sein mussten, ergab unter Verwendung der Trim und Fill Methode, dass in 30 (75%; REM) respektive 36 (90%; FEM) Meta-Analysen Studien fehlten. Die Höhe der mittleren Effektstärke änderte sich jedoch in keiner der Meta-Analysen durch die Korrektur signifikant. Der Rangkorrelationstest wurde ebenfalls eingesetzt und wurde in 14 Fällen (35%) auf dem 10%-Niveau signifikant.

Zusammenfassend lässt sich aus diesen Untersuchungen schließen, dass ausreichende Belege für die Existenz eines Publikationsbias in verschiedenen Forschungsbereichen vorliegen. In allen fünf Untersuchungen wurde ein Publikationsbias nachgewiesen. Das Ausmaß des Publikationsbias variiert in den fünf Studien jedoch, so dass es nicht möglich ist, die Ergebnisse dieser Untersuchungen auf andere Forschungsgebiete zu generalisieren (Sutton, 2005). Da ein Publikationsbias in jedem Forschungsfeld in unterschiedlichem Maße und mit unterschiedlich starken Auswirkungen auf die Höhe der mittleren Effektstärke auftreten kann, sollte seine Überprüfung für jeden Forschungsbereich separat durchgeführt werden. Aus den vorliegenden Ergebnissen kann nicht ohne weiteres auf das Vorliegen und die Höhe eines Publikationsbias im Bereich der Psychotherapieforschung geschlossen werden. Die aus einem Publikationsbias resultierende Überschätzung der mittleren Effektstärke kann geringen, mittleren oder starken Ausmaßes sein. So kann eine Publikationsverzerrung beispielsweise so groß sein, dass bei Berücksichtigung unveröffentlichter nicht-signifikanter Ergebnisse die Höhe der mittleren Effektstärke deutlich reduziert würde, zum Beispiel von einem großen Effekt zu einem kleinen, und somit die Wirksamkeit der untersuchten Intervention nach der Korrektur als deutlich niedriger eingeschätzt werden muss. Der Einfluss eines Publikationsbias kann aber auch so gering sein, dass sich die Höhe des mittleren Effektes trotz fehlender Studien mit Nullergebnissen nicht ändert und trotz des Vorliegens einer geringen Tendenz zur selektiven Veröffentlichung positiver Ergebnisse die Wirksamkeit einer Intervention substantiell gegeben bleibt (Rothstein et al., 2005). Deshalb sollte das Vorliegen eines Publikationsbias in jedem Forschungsbereich untersucht werden.

Im Bereich der Psychotherapieforschung wurde ein möglicher Publikationsbias bislang nicht umfassend geprüft. Zur Wirksamkeit psychotherapeutischer Interventionen bei psychischen Störungen wurden in den letzten Jahrzehnten zahlreiche Primärstudien sowie

darauf aufbauend Meta-Analysen zu deren Integration durchgeführt. Auch aus politischen Gründen ist in den letzten zwei Jahrzehnten die Orientierung an diesen empirischen Wirksamkeitsnachweisen und somit die evidenzbasierte Psychotherapie im Bereich der Leistungen der deutschen Gesundheitsversorgung relevant geworden (Fydrich & Schneider, 2007). Die Prüfung der Wissenschaftlichkeit nach bestimmten Evidenzkriterien des Wissenschaftlichen Beirats Psychotherapie ist im Psychotherapeutengesetz explizit als Voraussetzung für die Anerkennung einer Therapieform als Richtlinienverfahren verankert (Pulverich, 1999). Die Kosten für eine als wirksam erwiesene und deshalb als Richtlinienverfahren anerkannte Therapieform werden von den Krankenkassen als Heilbehandlung erstattet. Den empirischen Wirksamkeitsnachweisen kommt im Rahmen der evidenzbasierten Forschung, als Grundlage für die klinische Anwendung, also eine hohe praktische Relevanz zu. Die erbrachten Effektivitätsnachweise bieten eine Orientierung, welche therapeutischen Verfahren wirksam in der Praxis angewendet werden können, und dadurch auch für die Verteilung der Ressourcen in der Gesundheitsversorgung. Wenn, basierend auf überschätzten Wirksamkeitsnachweisen, unwirksame oder im Extremfall sogar schädliche Interventionen fälschlicherweise zur Anwendung in der Praxis empfohlen werden, würde dies zu unnötigem Leid bei den Patienten sowie vergeudeten Ressourcen im Gesundheitssystem führen (Song, Eastwood, Gilbody, Duley & Sutton, 2000). Aufgrund der Relevanz der Implementierung wirksamer Therapien bei psychischen Störungen ist die Überprüfung eines möglichen Publikationsbias zur Sicherung der Validität der Ergebnisse, indem einer Überschätzung des Behandlungseffektes sowie der Möglichkeit schwerwiegender Nebenwirkungen entgegengewirkt wird, unerlässlich.

Die vorliegende Arbeit widmete sich deshalb der Überprüfung eines möglichen Publikationsbias in Meta-Analysen zur Erforschung der Wirksamkeit psychotherapeutischer Interventionen in den drei Störungsbereichen Essstörungen, Schizophrenie und Depression. Des weiteren wurden Meta-Analysen, in denen die Wirksamkeit von Präventionsprogrammen bei Essstörungen und Depression untersucht wurde, eingeschlossen. Die Störungsgruppe der Essstörungen umfasst die Bulimia Nervosa (BN), die Anorexia Nervosa (AN) sowie die Binge Eating Störung (BED). Eine Zusatzkategorie des DSM-IV (APA, 1994) greift ein Vorliegen klinisch relevanter Essstörungssymptome auf, wenn nicht alle Kriterien einer einzelnen Essstörung vollständig erfüllt sind, und wird als “Eating disorders not otherwise specified“ (EDNOS) bezeichnet. Essstörungen haben hohe Prävalenzraten von 0.3-1% für BN (Legenbauer & Vocks, 2006), 0.3-3% für AN (Hoek, 2006) und 0.7-6.6 % für BED (Brownley, Berkman, Sedway, Lohr & Bulik, 2007; Vocks et al., 2010). AN ist

gekennzeichnet durch eine starke Angst vor einer Gewichtszunahme bei bestehendem Untergewicht, BN durch ausgeprägte Heisshungeranfälle und unangemessene Gegenmaßnahmen wie selbstinduziertes Erbrechen, während BED durch Heisshungeranfälle ohne anschließendes kompensatorisches Verhalten charakterisiert ist. Bei allen Essstörungen ist die Anwendung effektiver Therapieverfahren aufgrund der Schwere der Symptomatik und des damit verbundenen psychischen Leidensdrucks, bei häufigem Vorkommen in der Allgemeinbevölkerung, unerlässlich. Bei BN und BED zeigt sich die Kognitive Verhaltenstherapie (KVT; Jacobi, Paul & Thiel, 2004) als besonders wirksam zur Reduktion der Symptomatik, aber auch andere therapeutische Verfahren wie die Interpersonelle Therapie (IPT; Klerman, Weissman, Rounsaville & Chevron, 1984). Allerdings wurden die vorliegenden Wirksamkeitsnachweise nicht auf den Einfluss eines möglichen Publikationsbias hin überprüft. Für AN liegen bislang keine empirischen Wirksamkeitsnachweise auf meta-analytischer Ebene vor (Herpertz, Herpertz-Dahlmann, Fichter, Tuschen-Caffier & Zeeck, 2011).

Die Schizophrenie gehört zu den schwerwiegendsten psychiatrischen Erkrankungen. Die Angaben zur Lebenszeitprävalenz liegen zwischen 0.4% (McGrath, Saha, Chant & Welham, 2008) und 1% (Hahlweg & Dose, 1998). Die Erkrankung ist gekennzeichnet durch Positivsymptomatik wie Wahnvorstellungen sowie Negativsymptomatik wie affektive Verflachung oder Antriebsarmut und geht häufig mit weiteren schwerwiegenden Folgen einher, wie dem Verlust des Arbeitsplatzes sowie der sozialen Stellung. Häufig ist Psychoedukation, auch unter Einbezug der Angehörigen des Patienten, indiziert, insbesondere nach der ersten Diagnose der Erkrankung, um Informationen über die Symptomatik und zur Verbesserung der Medikamenten-Compliance zu vermitteln. Aber auch in der Rückfallprophylaxe kommen psychoedukative Familienprogramme zum Tragen (Hahlweg & Dose, 1998). Weiterhin gehören KVT und das „Integrierte psychologische Therapieprogramm“ (Roder, Brenner, Kienzle & Hodel, 1997), das auf eine Verbesserung der kognitiven und sozialen Fertigkeiten abzielt, zu den empfohlenen und am häufigsten eingesetzten Therapieverfahren bei Schizophrenie. Diese Therapieverfahren wurden empirisch auf ihre Wirksamkeit hin untersucht, sind jedoch ebenfalls nicht umfassend auf einen Publikationsbias geprüft worden.

Depression ist mit einer Lebenszeitprävalenz von 10-14% (Singleton, Bumpstead, O'Brien, Lee & Meltzer, 2001; Hautzinger, 1998) eine der am häufigsten vorkommenden psychischen Störungen und ist gekennzeichnet durch Antriebslosigkeit, Niedergeschlagenheit

und Interessenverlust, bis hin zu Suizidideen und akuter Suizidalität (Hautzinger, 1998). Aufgrund der Schwere der Symptomatik ist bei dieser Störung ebenfalls eine wirksame Therapie unerlässlich. Es liegt eine Vielzahl von Wirksamkeitsnachweisen für verschiedene therapeutische Verfahren vor, insbesondere für die KVT (Hautzinger, 1998; Lynch, Laws & McKenna, 2010), und weiterhin auch für die Interpersonelle Therapie (IPT; Klerman et al., 1984), die „Problemlöse-Therapie“ (Cujipers, van Straten & Warmerdam, 2007) und den psychoedukativen „Coping with Depression“-Kurs (Cuijpers, Munoz, Clarke & Lewinsohn, 2009), um einige zu nennen. Die Wirksamkeitsnachweise für psychotherapeutische Verfahren bei Depression wurden ebenfalls nicht umfassend auf einen möglichen Publikationsbias hin untersucht.

Für die vorliegende Arbeit wurden in einer umfassenden Literaturrecherche zunächst sämtliche bis September 2010 in deutscher oder englischer Sprache veröffentlichten Meta-Analysen zur Wirksamkeit psychotherapeutischer Interventionen bei Essstörungen, Schizophrenie und Depression sowie präventiver Maßnahmen bei Essstörungen und Depression gesichtet. Anschließend wurde geprüft, wie häufig unpublizierte Studien in diesen Meta-Analysen eingeschlossen waren, und wie häufig statistische Verfahren zur Überprüfung eines Publikationsbias eingesetzt wurden. Es zeigte sich, dass in nur 6 (35.29%) der 17 Meta-Analysen zu Essstörungen, in 4 (8.7%) der 46 Meta-Analysen zu Schizophrenie und in 16 (18.82%) der 85 Meta-Analysen zu Depression unpublizierte Studien eingeschlossen waren. Darüber hinaus wurden die vorgestellten Methoden zur Erfassung des Vorliegens eines möglichen Publikationsbias in nur sehr wenigen Meta-Analysen angewandt: in keiner der Meta-Analysen zu Essstörungen, in nur 1 (2.17%) Meta-Analyse zu Schizophrenie Eggers Regressionsanalyse (Egger et al., 1997), aber in keiner Begg und Mazumdars Rangkorrelationstest (1994) oder die Trim und Fill Methode (Taylor & Tweedie, 2000a, 2000b), und bei Depression in 6 (7.06%) Meta-Analysen die Regressionsanalyse, in 3 (3.53%) der Rangkorrelationstest, und in 12 (14.12%) Trim und Fill. In den drei Störungsbereichen ist damit ein nahezu vollständiges Fehlen statistischer Kontrolle eines Publikationsbias zu verzeichnen. Sowohl die bislang durchgeführte Inklusion unpublizierter Studien sowie die Überprüfung einer Publikations-Verzerrung in diesen Forschungsbereichen erweisen sich als unzureichend. Eine Absicherung der vorliegenden Wirksamkeitsnachweise gegen eine mögliche Überschätzung der Effektivität ist notwendig.

In der vorliegenden Untersuchung sollten aus diesem Grund sämtliche Meta-Analysen aus dem Bereich der Psychotherapieforschung zu Essstörungen, Schizophrenie und

Depression mittels des Rangkorrelationstests, der Regressionsanalyse sowie der Trim und Fill Methode auf einen Publikationsbias hin geprüft werden. Sämtliche Datensätze einer Meta-Analyse, die die folgenden Kriterien erfüllten, wurden in die Überprüfung eingeschlossen: 1) der mittlere gepoolte Effekt wurde berichtet; 2) die Primärstudien-Effektstärken oder die zu deren Kalkulation notwendigen Rohdaten waren angegeben; 3) ein Maß der Präzision jedes Primärstudieneffektes oder die zu dessen Kalkulation notwendigen Rohdaten waren angegeben (war das Effektstärkemaß ein Korrelationskoeffizient, so konnte stattdessen auch die Stichprobengröße verwendet werden); 4) die in einem notwendigen Zwischenschritt für die vorliegende Untersuchung nachberechneten mittleren Effektstärke-Schätzer mussten (bis auf geringe Abweichungen im Nachkommabereich) mit den in der Meta-Analyse angegebenen Effektstärken übereinstimmen; 5) der mittlere Effekt musste auf Homogenität geprüft worden sein und dieses Kriterium erfüllen; 6) der Datensatz musste mindestens sechs Studien umfassen; und 7) die Methode Trim und Fill wurde in der Meta-Analyse noch nicht angewandt. Konnten die notwendigen Daten einer Meta-Analyse nicht entnommen werden, so wurden die Autoren mit der Bitte um Bereitstellung der Daten angeschrieben. Sämtliche Berechnungen erfolgten mittels des Programms „Comprehensive Meta-Analysis“, Version 2.2 (Borenstein, Hedges, Higgins & Rothstein, 2005).

Die erste Untersuchung widmete sich den Meta-Analysen zu den einzelnen Störungsbildern der Essstörungen. In der zweiten Untersuchung wurden dieselben Methoden unter denselben Voraussetzungen auf Meta-Analysen zur Wirksamkeit von Psychotherapie bei Schizophrenie angewandt. In der letzten Untersuchung erfolgte dasselbe Vorgehen für das Störungsbild der Depression. Eine detaillierte Beschreibung des Aufbaus, der Methoden und Ergebnisse findet sich jeweils in Artikel 1, Artikel 2 und Artikel 3. Deshalb wird im Folgenden lediglich ein kurzer Überblick gegeben.

Untersuchung 1

In der ersten Untersuchung wurden Meta-Analysen zur Wirksamkeit psychotherapeutischer und präventiver Interventionen bei Essstörungen auf einen möglichen Publikationsbias hin geprüft. Eine umfassende Literaturrecherche, in der die Suchbegriffe „meta-analysis” oder “systematic review” und ”eating disorders” in Kombination verwendet wurden, wurde in den Datenbanken, Psyndex und PsyInfo durchgeführt. Zusätzlich wurde in den Literaturverzeichnissen einschlägiger Artikel und Bücher nach weiteren Meta-Analysen recherchiert. Die Meta-Analysen mussten vor September 2010 in deutscher oder englischer

Sprache veröffentlicht worden sein. Sämtliche therapeutischen und präventiven Interventionen wurden in die Überprüfung eingeschlossen. Meta-Analysen zur Wirksamkeit von Psychopharmaka wurden ausgeschlossen, da der Fokus der vorliegenden Untersuchung auf der Überprüfung psychotherapeutischer Interventionen lag. In den Meta-Analysen mussten die Störungsbilder BN, AN, BED oder EDNOS untersucht worden sein, diagnostiziert anhand des DSM-III (APA, 1980), DSM-III-R (APA, 1987), DSM-IV (APA, 1994), DSM-IV-TR (APA, 2000), ICD-9 (Degkwitz, Helmchen, Kockott & Mombour, 1980) oder des ICD-10 (World Health Organization, 1992), oder sie mussten gesunde sowie Risikostichproben umfassen, an denen Interventionen zur Prävention von Essstörungen durchgeführt wurden. Meta-Analysen, in denen die Wirksamkeit von Interventionen nicht ausschließlich bei Essstörungen untersucht wurde, sondern in die zusätzlich auch Studien mit anderen Störungsbildern eingingen, wurden ausgeschlossen, wenn keine differenziellen Wirksamkeitsnachweise für Essstörungen vorlagen.

Zwei Meta-Analysen (Hay, Bacaltchuk, Stefano & Kashyap, 2009; Stice & Shaw, 2004) erfüllten sämtliche Kriterien und konnten in die Untersuchung eingeschlossen werden. In der Meta-Analyse von Hay et al. (2009) wurde die Wirksamkeit von KVT sowie von verschiedenen Therapieverfahren, die zur Effektstärkenberechnung in einer Gruppe zusammengefasst wurden, untersucht. Die Gruppe der Therapieverfahren umfasste IPT, hypno-behaviorale Therapie, dialektisch-behaviorale Therapie, unterstützende Therapie und weitere Therapieverfahren, die behaviorale Interventionen zur Gewichtsreduktion beinhalteten. BN, BED und EDNOS wurden in dieser Meta-Analyse zu einer Störungsgruppe zusammengefasst. Dieser Meta-Analyse konnten elf Datensätze entnommen werden. In der Meta-Analyse von Stice und Shaw (2004) wurde die Wirksamkeit von Präventionsprogrammen bei allgemeinen und Risikostichproben untersucht. Drei Datensätze dieser Meta-Analyse erfüllten die Einschlußkriterien.

Die drei dargestellten Methoden, Begg und Mazumdars Rangkorrelationstest (1994), Eggers Regressionsanalyse (Egger et al., 1997) und Taylor und Tweedies Trim und Fill Methode (2000a, 2000b), wurden zur Überprüfung eines Publikationsbias angewandt. Das Ergebnis des Rangkorrelationstests fiel bei 2 Datensätzen (18.2%) unter der zweiseitigen Signifikanztestung sowie bei 3 Datensätzen (27.3%) unter der einseitigen Testung signifikant aus. Eggers Regressionsanalyse ergab bei denselben 3 Datensätzen (27.3%) einen signifikanten Publikationsbias, sowohl unter der ein- wie auch der zweiseitigen Prüfung. Betroffen waren die Effektivität von KVT zur Erhöhung von *Remissionsraten*, sowie die

Wirksamkeit von Präventionsprogrammen zur Verringerung der Risikofaktorausprägungen *Idealisierung von Schlankheit* und *negativer Affekt*. Die Überprüfung mittels Trim und Fill ergab, daß in 9 Datensätzen (64.3%) aufgrund von Asymmetrie fehlende Studien eingefügt werden mussten. Allerdings fiel keine der Reduktionen der mittleren Effektstärken signifikant aus, da sämtliche Originaleffektstärken innerhalb der Konfidenzintervalle ihrer korrigierten Effektstärken lagen. Zusätzlich verblieben vier Effektstärken, die sich vor der Korrektur signifikant von null unterschieden, ebenfalls nach dem Einfügen fehlender Studien signifikant, und weiterhin vier zuvor nicht-signifikante Effektstärken auch nach der Korrektur nicht-signifikant. Die mittlere Effektstärke des neunten Datensatzes, in dem zwei wirksame Therapieverfahren, KVT und die Gruppe der Psychotherapien, hinsichtlich des Outcome-Maßes Erhöhung der *Remissionsraten* miteinander verglichen wurden, fiel vor der Korrektur nicht-signifikant aus. Nach der Korrektur und dem Einfügen zweier fehlender Studien zeigte dieser Vergleich jedoch eine signifikant höhere Wirksamkeit der KVT an. Insgesamt fügte Trim und Fill in die Datensätze bis zu vier fehlende Studien ein. Diese Maximalanzahl betraf einen Datensatz zur Wirksamkeit von Präventionsprogrammen bezüglich der Reduktion von *negativem Affekt* zum Katamnesezeitpunkt (Stice & Shaw, 2004). Die drei Datensätze, bei denen der Rangkorrelationstest sowie die Regressionsanalyse signifikant ausfielen, sollten einerseits mit einer gewissen Vorsicht betrachtet werden. Da andererseits aber das Einfügen fehlender Studien durch die Trim und Fill Methode in diesen drei Datensätzen zu keiner signifikanten Reduktion der mittleren Effektstärken führte, ist der Einfluss des Publikationsbias auf die Höhe der mittleren Effektstärken gering. Die Wirksamkeit der untersuchten Methoden zur Veränderung der genannten Outcome-Maße bleibt substantiell gegeben. Bei insgesamt 11 der 14 Datensätze fiel keine der 3 Methoden signifikant aus. Aus diesen Ergebnissen lässt sich schließen, dass KVT und die Gruppe psychotherapeutischer Verfahren sowie Programme zur Prävention von Essstörungen, trotz einer geringen Tendenz zur selektiven Publikation positiver Ergebnisse, als wirksam eingeschätzt werden können.

Untersuchung 2

In der zweiten Untersuchung wurde dasselbe methodische Vorgehen wie in der ersten Untersuchung auf Meta-Analysen, in denen die Wirksamkeit therapeutischer Interventionen bei Schizophrenie untersucht wurde, angewandt. Die Literaturrecherche wurde anhand der Suchbegriffe „meta-analysis“ oder „systematic review“ und „schizophrenia“, „schizoaffective disorder“ oder „delusional disorder“ durchgeführt. Die Diagnosestellung musste in den Primärstudien ebenfalls anhand des DSM-III (APA, 1980), DSM-III-R (APA, 1987), DSM-

IV (APA, 1994), DSM-IV-TR (APA, 2000), ICD-9 (Degkwitz et al., 1980) oder des ICD-10 (World Health Organization, 1992) vorgenommen worden sein. Die Einschlusskriterien, bezogen auf die Störung Schizophrenie anstelle von Essstörungen, waren mit denen der ersten Untersuchung identisch.

Zehn Meta-Analysen konnten eingeschlossen werden. In diesen zehn Meta-Analysen wurden folgende Therapieverfahren untersucht: KVT wurde in den Meta-Analysen von Jones, Cormac, Silveira da Mota Neto und Campbell (2010), Lincoln, Suttner und Nestoriuc (2008), Lynch et al. (2010), Wykes, Steel, Everitt und Tarrier (2008) sowie Zimmermann, Favrod, Trieu und Pomini (2005) auf seine Wirksamkeit überprüft. Psychosoziale Familieninterventionen wurden in der Meta-Analyse von Pharoah, Mari, Rathbone und Wong (2010) untersucht. Familientherapie kombiniert mit Psychoedukation wurde in drei Meta-Analysen von Pitschel-Walz (1997), Pitschel-Walz und Engel (1997) sowie Pitschel-Walz, Leucht, Bäuml, Kissling und Engel (2001) untersucht. In der Meta-Analyse von Pitschel-Walz (1997) wurde zusätzlich in einem Datensatz ausschließlich die Effektivität von Psychoedukation überprüft. In der Meta-Analyse von Bola (2006) wurde die Wirksamkeit von Psychotherapie, kombiniert mit medikamentöser Behandlung, verglichen mit Psychotherapie mit medikationsfreien Intervallen. Insgesamt erfüllten 22 Datensätze dieser 10 Meta-Analysen die Einschlusskriterien der vorliegenden Untersuchung.

Der Rangkorrelationstest (Begg & Mazumdar, 1994) wurde bei 2 (9.09%) der Datensätze signifikant (ein- und zweiseitige Testung). Die Regressionsanalyse (Egger et al., 1997) fiel unter der einseitigen Testung bei 4 (18.18%) Datensätzen signifikant aus, und unter der zweiseitigen Testung bei 3 (13.64%) Datensätzen. In zwei dieser Datensätze wurde die Wirksamkeit von KVT zur Reduktion von *Positivsymptomatik,* in einem weiteren die Wirksamkeit von KVT zur Reduktion *schizophrener Gesamtsymptomatik*, und im vierten Datensatz von familien-orientierten Interventionen zur Verringerung der *Anzahl stationärer Einweisungen* untersucht. Diese Effektstärken sollten einerseits mit einer gewissen Vorsicht betrachtet werden. Die Trim und Fill Methode fand jedoch bei 2 dieser Datensätze keine fehlenden Studien, und fügte bei den weiteren 2 Datensätzen fehlende Studien ein, ohne dass die mittleren Effektstärken signifikant reduziert wurden, sodass alle 4 Effektstärken dennoch als relativ robust gegen den Einfluss eines Publikationsbias betrachtet werden können. Die Wirksamkeit der Interventionen blieb signifikant gegeben.

Trim und Fill (Taylor & Tweedie, 2000a, 2000b) fügte insgesamt bei 15 der 22 Datensätze fehlende Studien ein. Erwartungsgemäß wurden alle mittleren Effektstärken durch

die Korrektur reduziert. Allerdings fiel keine der Reduktionen signifikant aus. Weiterhin blieben alle 12 der 15 mittleren Effektstärken, die sich vor der Korrektur zufallskritisch von Null unterschieden, auch nach der Reduktion signifikant. Zwei Effektstärken fielen sowohl vor als auch nach der Korrektur nicht-signifikant aus: KVT erwies sich als nicht wirksam zur Reduktion der *Rückfallraten*, und psychosoziale Familieninterventionen als unwirksam bezüglich einer *Verringerung der Therapieabbruch-Rate (Zeitraum 7-12 Monate)*. Für eine weitere Effektstärke änderte sich jedoch das Signifikanzmuster: sie fiel vor der Korrektur nur knapp nicht-signifikant aus, wurde danach aber signifikant. Hier fehlte entgegen der Erwartung eine Studie mit einem Effekt zugunsten der Therapiebedingung, psychosoziale Familieninterventionen, bezogen auf das Outcome-Maß *Verringerung der Therapieabbruch-Rate (Zeitraum 13-24 Monate)*. Für Familieninterventionen änderte sich somit, bezogen auf dieses Outcome-Maß, unter Berücksichtigung des Einfluss eines Publikationsbias die Wirksamkeitseinschätzung.

Die maximale Anzahl fehlender Studien, die Trim und Fill in einen Datensatz einfügte, betrug fünf. Diese maximale Anzahl betraf einen Datensatz, in dem die Wirksamkeit von KVT zur Reduktion von *Positivsymptomatik* untersucht wurde. Insgesamt fand keine der 3 Methoden in 18 der 22 Datensätze signifikante Hinweise auf einen Publikationsbias, da die Veränderung bei keiner der durch Trim und Fill reduzierten Effektstärken signifikant ausfiel. Es besteht somit eine hohe Übereinstimmung zwischen den drei angewendeten Verfahren. Nur für psychosoziale Familieninterventionen änderte sich die Wirksamkeitseinschätzung, und zwar dahingehend, dass der Effekt zuvor knapp die Signifikanz verfehlt hatte und unter Hinzufügen einer fehlenden Studie signifikant wurde. Für alle Interventionen außer psychosozialen Familieninterventionen lässt sich schlussfolgern, dass die Wirksamkeitseinschätzung trotz einer geringen Tendenz zur selektiven Publikation signifikanter Ergebnisse nicht substantiell verändert wird. Auch für das Störungsbild der Schizophrenie kann deshalb das Fazit gezogen werden, dass die Effektstärken der untersuchten Interventionen nur in geringem Maße durch Publikationsbias verzerrt sind.

Untersuchung 3

Die dritte Untersuchung folgte ebenfalls dem Aufbau der beiden vorausgehenden Studien. In dieser Analyse wurden die Methoden zur Überprüfung eines Publikationsbias auf Meta-Analysen angewandt, in denen die Wirksamkeit therapeutischer und präventiver Interventionen bei Depression untersucht wurde. Es wurden dieselben Einschlusskriterien

verwendet, und als Suchbegriffe dienten wiederum „meta-analysis“ oder „systematic review“, die für die Recherche nach Meta-Analysen zum Störungsbild Depression mit einem der Begriffe „depression“, „major depression“, „depressive disorder“ oder „dysthymia“ kombiniert wurden. Die Diagnose der Patienten musste in den Primärstudien anhand derselben Diagnosesysteme, wie für die erste und zweite Untersuchung dargestellt, vorgenommen worden sein. Zusätzlich waren jedoch auch Stichproben erlaubt, in denen die Patienten auf einer Selbstbeurteilungsskala über dem Cut-Off-Wert für eine auffällige Depressionssymptomatik lagen. Für den Einschluss von Meta-Analysen zur Prävention von Depression oder depressiven Rezidiven galt, dass die Studienteilnehmer keine akute depressive Episode aufweisen durften, und ein Instrument zur Veränderungsmessung depressiver Symptomatik eingesetzt sowie die Wirksamkeit einer präventiven Intervention zur Verhinderung einer Depression untersucht worden sein musste.

Für das Störungsbild Depression lag bereits eine umfassendere Analyse zum Publikationsbias vor (Cuijpers et al., 2010). In dieser explizit zur Überprüfung eines Publikationsbias durchgeführten Meta-Analyse mit 117 randomisierten kontrollierten Primärstudien wurden ebenfalls Begg und Mazumdars Rangkorrelationsmethode (1994), Eggers Regressionsanalyse (Egger et al., 1997) und die Trim und Fill Methode (Taylor & Tweedie, 2000a, 2000b) angewendet. In 37 der 43 Datensätze dieser Meta-Analyse wurde ein Publikationsbias gefunden, und die Anwendung der Trim und Fill Methode führte bei allen betroffenen mittleren Effektstärken zu starken Reduktionen. Der mittlere Effekt, der über sämtliche eingeschlossenen Studien berechnet wurde, musste von einem Hedges $g = 0.67$ auf $g = 0.42$ reduziert werden, wobei 51 fehlende Studien eingefügt wurden. Der Rangkorrelationstest sowie die Regressionsanalyse wurden ebenfalls bei fast allen Datensätzen signifikant. Gleichzeitig fiel aber auch die Heterogenität in 40 Datensätzen dieser Meta-Analyse hochsignifikant aus. Da die auf einer Homogenitätsannahme beruhenden Methoden zur Überprüfung des Publikationsbias bei Vorliegen von Heterogenität jedoch nicht angewendet werden dürfen, liegt hier eine Verletzung der statistischen Voraussetzungen vor. Die Autoren schlussfolgern dennoch, dass die Wirksamkeit therapeutischer Interventionen bei Depression aufgrund eines Publikationsbias stark überschätzt wird. Anhand der Anwendung der Methoden bei ausschließlich homogenen Datensätzen sollte deshalb das Risiko eines Publikationsbias für den Störungsbereich Depression in der vorliegenden Untersuchung erneut geprüft werden, um hier gleichzeitig einer Überschätzung des Ausmaßes eines Publikationsbias entgegenzuwirken.

Neunzehn Meta-Analysen erfüllten sämtliche Einschlusskriterien und wurden in die Untersuchung aufgenommen (Barbato & D´Avanzo, 2008; Beltman, Voshaar & Speckens, 2010; Bortolotti, Menchetti, Bellini, Montaguti & Berardi, 2008; Cuijpers, Brannmark & van Straten, 2008; Cuijpers et al., 2009; Cuijpers et al., 2007; Cuijpers, Van Straten, Smits & Smit, 2006a; Cuijpers, Van Straten & Smit, 2006b; De Maat, Dekker, Schoevers & De Jonghe, 2006, 2007; Dennis, 2005; Ekers, Richards & Gilbody, 2008; Haby, Donnelli, Corry & Vos, 2006; Himelhoch, Medoff & Oyeniyi, 2007; Lynch et al., 2010; Neumeyer-Gromen, Lampert, Stark & Kallischnigg, 2004; Newton-Howes, Tyrer & Johnson, 2006; Pampallona, Bollini, Tibaldi, Kupelnick & Munizza, 2004; Reinecke, Ryan & DuBois, 1998). In diesen Meta-Analysen wurde die Wirksamkeit von KVT, Verhaltenstherapie, mehreren psychotherapeutischen Verfahren zu einer Gruppe zusammengefasst, einer Kombinationsbehandlung aus Psychotherapie und Psychopharmaka, des „Coping with Depression"- Psychoedukationskurses, Paartherapie oder präventiven Interventionen untersucht. Diesen 19 Meta-Analysen konnten insgesamt 31 Datensätze entnommen werden.

Begg und Mazumdars Rangkorrelationstest (1994) und Eggers Regressionsanalyse (Egger et al., 1997) zeigten in je 9 (29.03%) Datensätzen unter einseitiger Signifikanztestung einen Publikationsbias an, weiterhin in 5 (16.13%) respektive 6 (19.35%) derselben Datensätze unter zweiseitiger Testung. In diesen Datensätzen wurden die Wirksamkeit von KVT, Verhaltenstherapie, Gruppentherapie sowie des „Coping with Depression"- Kurses zur Reduktion *depressiver Symptomatik*, die Wirksamkeit von Psychotherapie zur Reduktion *depressiver Symptomatik* sowie zur Erhöhung der *Remissionsraten*, und von Psychotherapie zur Reduktion *depressiver Symptomatik* bei depressiven Patienten mit komorbider Persönlichkeitsstörung untersucht. Die Effektstärken dieser Datensätze sollten einerseits mit Vorsicht betrachtet werden. In zwei dieser Datensätze wurden durch die Trim und Fill Methode (Taylor & Tweedie, 2000a, 2000b) andererseits aber keine fehlenden Studien eingefügt, sodass deren Effektschätzer dennoch als relativ robust betrachtet werden können. In weitere sieben dieser Datensätze wurden Studien eingefügt, aber keine der Reduktion der mittleren Effektstärken fiel signifikant aus. Für zwei der untersuchten Interventionen änderte sich jedoch die Wirksamkeitseinschätzung, beide zuvor signifikanten Effektstärken fielen nach der Reduktion nicht mehr signifikant aus. Im ersten Datensatz wurde die Wirksamkeit von Psychotherapie bei Patienten mit und ohne komorbide Persönlichkeitsstörung hinsichtlich der Reduktion *depressiver Symptomatik* verglichen. Die nach der Korrektur um einen Publikationsbias nicht mehr signifikante mittlere Effektstärke zeigte, dass die Behandlung für beide Patientengruppen gleich wirksam ist. Im zweiten Datensatz wurde die Wirksamkeit von

Psychotherapie versus Psychopharmaka hinsichtlich der Reduktion von *Dropoutraten* verglichen, deren Effektivität sich ebenfalls nach der Korrektur nicht mehr unterschied.

Insgesamt ergab die Überprüfung mittels Trim und Fill in 12 (38.71%) der 31 Datensätze fehlende Studien. Die Höchstanzahl von acht eingefügten Studien betraf einen Datensatz zur Wirksamkeit des „Coping with Depression"- Kurses bezüglich der Reduktion *depressiver Symptomatik*. Insgesamt führte aber keine der Imputationen fehlender Studien zu einer signifikanten Veränderung der korrigierten gegenüber der alten Effektstärke. Weiterhin verblieben acht der Effektstärken, die vor der Überprüfung signifikant waren, danach ebenfalls signifikant. Zwei weitere Effektstärken fielen vor der Korrektur bereits nicht-signifikant aus, da jeweils zwei wirksame Behandlungsmethoden miteinander verglichen wurden, und verblieben dieses auch danach. Die zwei Datensätze, bei denen sich die Wirksamkeitseinschätzung der darin untersuchten Interventionen durch den Publikationsbias änderte, wurden bereits dargestellt.

Es zeigte sich eine hohe Konkordanz der drei verwendeten Methoden: In 21 der 31 geprüften Datensätze verwies keine der Methoden auf einen signifikanten Publikationsbias. Für nur zwei Interventionen änderte sich unter dem Einfluss des Publikationsbias die Wirksamkeitseinschätzung. Mit Hinblick auf die Anwendung dieser Interventionen sollten die neuen, bias-korrigierten Effektstärkeschätzer als Orientierung für die klinische Praxis dienen. Für alle anderen psychotherapeutischen und präventiven Interventionen für das Störungsbild Depression verbleibt die Effektivität unverändert und substantiell. In Abgrenzung zu der Untersuchung von Cuijpers et al. (2010) lässt sich schlussfolgern, dass eine deutlich geringere Tendenz zur selektiven Publikation positiver Ergebnisse vorliegt. Die Ergebnisse der Arbeit von Cuijpers et al. gehen mit hoher Wahrscheinlichkeit auf die Heterogenität in den von ihm untersuchten Datensätzen zurück und stellen damit falsch positive Treffer dar.

Allgemeine Diskussion

In den drei dargestellten Untersuchungen konnte jeweils ein Teilbereich der Psychotherapieforschung, zur Wirksamkeit therapeutischer Interventionen bei Essstörungen, Schizophrenie und Depression sowie präventiver Interventionen zur Verhinderung der Entwicklung von Essstörungen und Depression, auf einen Publikationsbias hin überprüft werden. Die Untersuchungen waren vom methodischen Vorgehen her vergleichbar aufgebaut, und in allen Bereichen zeigte sich, dass ein Publikationsbias in einem eher geringen Ausmaß vorlag. Die Befunde der vorliegenden Arbeit sprechen mehrheitlich dafür, dass die

Wirksamkeitsnachweise für Psychotherapie und Präventionsmaßnahmen bei Essstörungen und Depression sowie für die Therapie der Schizophrenie valide und robust gegen den Einfluss eines Publikationsbias sind, und die Effektstärken größtenteils nicht überschätzt werden. Die Rangkorrelationsmethode wurde unter zweiseitiger Testung bei 2 (18.2%) der Datensätze aus dem Bereich der Essstörungen signifikant, ebenfalls bei 2 (9.09%) der Datensätze, die den Meta-Analysen zur Schizophrenie entnommen wurden, und bei 5 (16.13%) der Datensätze zur Depression. Die Regressionsanalyse fiel bei 3 (27.3%) der Datensätze aus dem Bereich der Essstörungen, bei 3 (13.64%) Datensätzen der Meta-Analysen zur Schizophrenie, und bei 6 (19.35%) der Datensätze zur Depression signifikant aus, ebenfalls unter zweiseitiger Testung. Die Prozentangaben in den 3 Untersuchungen liegen somit zwischen 9 und 27%, was ein eher geringes Ausmaß eines Publikationsbias indiziert. Weiterhin führte das gelegentliche Einfügen fehlender Studien mittels der Trim und Fill Methode in diesen von Beggs und Eggers Methoden ausgewiesenen Datensätzen niemals zu einer signifikanten Veränderung der mittleren Originaleffektstärke gegenüber der korrigierten Effektstärke, und auch in keinem der übrigen untersuchten Datensätze wurde eine der Effektstärkenreduktionen mittels Trim und Fill signifikant. Zusammenfassend liegt damit das positive Ergebnis vor, dass ein nur geringes Ausmaß selektiver Publikation gegeben ist. Insgesamt änderte sich nach der Korrektur um den Einfluss des Publikationsbias bei vier Datensätzen die Wirksamkeitseinschätzung der darin untersuchten Interventionen. Zwei der Effektstärken fielen nach der Korrektur signifikant aus, und 2 weitere wurden nicht-signifikant. KVT ist anderen Psychotherapieformen bei Essstörungen zur Erhöhung der Remissionsrate überlegen, psychosoziale Familieninterventionen sind zur Verringerung von vorzeitigem Therapieabbruch bei Schizophrenen wirksamer als eine Standardbehandlung, Psychotherapie ist unter Berücksichtigung des Publikationsbias bei Patienten mit komorbider Persönlichkeitsstörung ebenso wirksam wie bei Patienten mit reiner Depression, und Psychotherapie zur Reduktion der Dropoutraten bei Depressiven ebenso wirksam wie Psychopharmaka. Da die nicht-signifikanten Effektstärken nicht anzeigen, dass die betreffenden Interventionen nicht mehr wirksam sind, sondern dass in einem Fall eine Intervention bei zwei verschiedenen Stichproben gleich wirksam ist, und im anderen Fall zwei wirksame Behandlungsmethoden einander ebenbürtig sind, wird hier keine Intervention durch den Einfluss eines Publikationsbias unwirksam. Weiterhin fielen insgesamt sieben Effektschätzer aus Vergleichen von Behandlungsmethoden mit Kontrollgruppen vor und nach der Korrektur nicht-signifikant aus, sodass festgestellt werden muss, dass nicht alle der in den überprüften Meta-Analysen untersuchten Interventionen als wirksam gelten können. Diese

Interventionen beziehen sich auf die Störungsbilder Essstörungen und Schizophrenie. Alle Interventionen für Depression sind hingegen als wirksam einzuordnen.

Als wichtige Schlussfolgerungen für die Anwendung der Interventionen in der klinischen Praxis lässt sich aus diesen Ergebnissen ableiten, dass sich erstens die Nicht-Wirksamkeitseinschätzung von sieben Interventionen unter dem Einfluss eines Publikationsbias nicht änderte. Diese Interventionen können zur Reduktion der betreffenden Outcome-Maße für die Störungsbilder der Essstörungen sowie Schizophrenie nicht empfohlen werden. Zweitens wurden die Interventionen identifiziert, die robust gegen einen Publikationsbias und wirksam zur Behandlung von Essstörung, Schizophrenie und Depression, sowie zur Prävention von Essstörungen und Depression sind. Drittens konnten die durch den Einfluss eines Publikationsbias in ihrer Wirksamkeit veränderten Interventionen bestimmt werden, und die neu berechneten Effektstärkeschätzer sollten als Orientierung für die praktische Anwendung der betreffenden Verfahren dienen. Die nach der Überprüfung auf einen Publikationsbias als signifikant ausgewiesenen Effektstärken können im Rahmen der evidenzbasierten Psychotherapie als robuste und verlässliche Orientierung für die Wahl der wirksamsten Interventionen bei Essstörungen, Schizophrenie und Depression und zu deren Prävention dienen.

Die vorliegende Arbeit schließt damit eine Lücke in der bisherigen Wirksamkeitsforschung und in der Anwendung meta-analytischer Methoden. Da in den bislang veröffentlichten Meta-Analysen zu den untersuchten Störungsbildern kaum unpublizierte Studien eingeschlossen waren und ein Publikationsbias nur in einer äußerst geringen Anzahl der Meta-Analysen untersucht wurde, blieb sein Ausmaß bislang unbekannt, und es war nicht sicher, ob die Effektivität therapeutischer Interventionen überschätzt wird.

Die Ergebnisse für das Störungsbild Depression sind im Vergleich zu der Untersuchung von Cuijpers et al. (2010) zu diskutieren. Cuijpers et al. (2010) fanden unter Anwendung des Rangkorrelationstests, der Regressionsanalyse und der Trim und Fill Methode ein sehr hohes Ausmaß an Publikationsbias. Dieses Ergebnis kam jedoch unter Nichtbeachtung der Homogenitätsannahme, eine der statistischen Voraussetzungen der angewandten Methoden, zustande und muss deshalb in Zweifel gezogen werden. Die dritte Untersuchung der vorliegenden Arbeit zielte aus diesem Grunde zusätzlich darauf ab, die Ergebnisse von Cuijpers et al. (2010) an homogenen Datensätzen zu überprüfen. Es zeigte sich, dass unter Beachtung der Homogenitätsannahme ein wesentlich geringeres Ausmaß an Publikationsbias nachzuweisen ist: Der Rangkorrelationstest und die Regressionsanalyse

wiesen bei nur wenigen der 31 geprüften Datensätze einen Publikationsbias nach, während das Einfügen fehlender Studien durch Trim und Fill bei keinem Datensatz zu einer signifikanten Reduktion der mittleren Effektstärke führte, und nur zwei Effektstärken nach der Überprüfung nicht-signifikant ausfielen. Unsere Untersuchung lässt deshalb den Schluss zu, dass die hohen Fallzahlen von Cuijpers et al. (2010) als falsch positive Treffer zu werten sind und das Ausmaß eines Publikationsbias im Bereich der Depressionsforschung deutlich niedriger anzusiedeln ist. Da in der vorliegenden Studie aufgrund der Heterogenität allerdings nicht dieselben Datensätze wie in der Untersuchung von Cuijpers et al. (2010) geprüft wurden, weist die unserer Untersuchung zugrundeliegende Stichprobe von Studien nur teilweise Überschneidungen mit der von Cuijpers et al. auf. Dennoch ist es unwahrscheinlich, dass die große Differenz zwischen den Ergebnissen beider Überprüfungen auf die Unterschiede zwischen den zugrunde liegenden Studien zurückgeht. In beiden Untersuchungen wurden hauptsächlich publizierte Primärstudien eingeschlossen, und zu einem großen Teil wurde darin die Wirksamkeit derselben Therapieverfahren untersucht. Aufgrund dessen ist es wahrscheinlich, dass die differierenden Ergebnisse auf die Berücksichtigung der statistischen Voraussetzungen der angewandten Methoden in der vorliegenden Arbeit zurückzuführen sind.

Die vorliegenden Ergebnisse sind mit denen der Untersuchungen von Egger et al. (1997), Sutton et al. (2000), Sterne et al. (2000) und Jennions und Moller (2002) vergleichbar. Auch in diesen Studien waren das Ausmaß sowie die Auswirkungen eines Publikationsbias auf die Höhe der mittleren Effektstärken nicht gravierend. Die Ergebnisse des Rangkorrelationstests und der Regressionsanalyse lagen in diesen Untersuchungen zwischen 10 bis 38% von einem Publikationsbias betroffener Datensätze, und dieses in den Untersuchungen von Egger et al. (1997), Sterne et al. (2000) sowie Jennions und Moller (2002) sogar bei einem liberaleren Signifikanzniveau von 10%. Obwohl die Ergebnisse zwischen den Untersuchungen in einem gewissen Maße variieren, liegt dennoch in keinem Fall ein hohes Ausmaß an Publikationsbias vor. Song et al. (2002), Sutton et al. (2000), Jennions und Moller (2002) fanden weiterhin unter Verwendung der Trim und Fill Methode nur in wenigen Fällen starke Veränderungen der mittleren Effektstärken, was ebenfalls mit den Ergebnissen der vorliegenden Untersuchungen vergleichbar ist. Deshalb bleibt zu hoffen, dass ein Publikationsbias in den meisten Forschungsgebieten zwar in einem geringen, aber eher selten in einem massiven Ausmaß vorliegt. Dennoch sollte ein Publikationsbias immer kontrolliert werden, da nicht sicher ist, ob die Ergebnisse dieser Untersuchungen ohne weiteres auf andere Bereiche generalisiert werden können. Song et al. (2002) fanden

beispielsweise ein wesentlich höheres Ausmaß an Publikationsbias mittels des Rangkorrelationstests: Dieser wurde bei 82% der von ihnen überprüften Meta-Analysen auf einem 5%-Niveau signifikant.

Einige wichtige Limitationen der vorliegenden Arbeit sind zu diskutieren. Für jedes Störungsbild erfüllte nur ein kleiner Teil der Meta-Analysen die Einschlusskriterien. Im Bereich der Essstörungen konnten lediglich 2 (11.76%) der 17 publizierten Meta-Analysen in die Untersuchung eingeschlossen werden, bei der Schizophrenie 10 (21.74%) von 46 Meta-Analysen und bei der Depression 19 (22.35%) der insgesamt 85 Meta-Analysen. In einem Großteil der ausgeschlossenen Meta-Analysen waren die Effektstärken und Präzisionsmaße der Primärstudien nicht angegeben, und die betreffenden Autoren stellten die Daten auch auf Nachfrage nicht zur Verfügung. Es ist nicht sicher, ob die vorliegenden Ergebnisse auf die ausgeschlossenen Meta-Analysen und somit den gesamten Forschungsstand zu Essstörungen, Schizophrenie und Depression generalisiert werden können. Die statistische Überprüfung eines Publikationsbias sollte deshalb als Routineverfahren in jeder Meta-Analyse durchgeführt werden (Sutton, 2005). In den Richtlinien der American Psychiatric Association (APA) zur Durchführung und Veröffentlichung von Meta-Analysen, den „Meta-Analysis Reporting Standards“ (MARS; APA, 2008, 2009) sowie in den von einer Expertengruppe erarbeiteten Richtlinien „Preferred reporting items for systematic reviews and meta-analyses” (Prisma; Moher, Liberati, Tetzlaff & Altmann, 2009) wird die Überprüfung eines Publikationsbias in jeder Meta-Analyse ausdrücklich empfohlen. Neben der unmittelbar gewährleisteten Überprüfung würde damit das Problem, dass notwendige Daten für eine nachträgliche Berechnung nicht verfügbar sind und Meta-Analysen von nachfolgenden Prüfungen ausgeschlossen werden müssen, ebenfalls gelöst.

In diesen Richtlinien wird weiterhin gefordert, dass für jede Primärstudie einer Meta-Analyse die Effektstärken sowie die Stichprobengröße (MARS) oder die Effektstärken zusammen mit der Angabe ihres Konfidenzintervalls als Präzisionsmaß (Prisma) berichtet werden. Diese sollten zumindest dann in Meta-Analysen angegeben werden, wenn eine Überprüfung des Publikationsbias darin nicht durchgeführt wurde, da Meta-Analysen bei Zugänglichkeit der Daten zumindest nachträglich überprüft werden können. In zukünftigen Meta-Analysen sollten deshalb erstens unpublizierte Studien eingeschlossen werden, zweitens die Primärstudienstatistiken angegeben werden, und drittens im günstigsten Falle statistische Überprüfungen eines Publikationsbias anhand der auch in der vorliegenden Arbeit angewendeten Methoden durchgeführt werden. Abschließend ist zu darauf hinzuweisen, dass

auch die Heterogenität der eingeschlossenen Daten immer geprüft werden sollte (APA, 2008; 2009), und bei Vorliegen durch Moderator-Analysen gegebenenfalls aufzuklären ist (Borenstein et al., 2009; Rustenbach, 2003). Dieses ist für die Prüfung eines Publikationsbias relevant, da die Methoden nur bei homogenen Datensätzen angewandt werden dürfen. Die vorliegende Arbeit soll deshalb über das positive Ergebnis eines nur geringen Ausmaßes an Publikationsbias hinaus dazu dienen, ein gesteigertes Problembewußtsein für das mögliche Vorliegen eines Publikationsbias sowie für die Voraussetzungen und die Notwendigkeit seiner Überprüfung zu schaffen.

Weiterhin könnte die Repräsentativität der vorliegenden Untersuchung hinsichtlich der Generalisierbarkeit der Ergebnisse auf die nicht überprüften Meta-Analysen dadurch eingeschränkt sein, dass die Meta-Analysen, die die Daten vollständig angegeben hatten und homogene Datensätze berichteten, möglicherweise sorgfältiger durchgeführt wurden und deshalb ein geringeres Ausmaß an Publikationsbias aufweisen, oder dass es andere zugrunde liegende Faktoren gibt, durch die das Ergebnis der vorliegenden Prüfung positiver ausfällt, als es für die übrigen Meta-Analysen der Fall sein könnte. Auch deshalb läßt sich das positive Ergebnis der vorliegenden Arbeit möglicherweise nur mit Einschränkungen auf den gesamten Forschungsbereich zu den untersuchten Störungen generalisieren und sollte sicherheitshalber nur auf die hier geprüften Meta-Analysen bezogen werden.

Leider waren in der vorliegenden Arbeit nicht alle mittleren Effektstärken sowie Konfidenzintervalle exakt replizierbar. Im Bereich der Nachkommastellen liegen einige Abweichungen der nachberechneten Werte von den Originaldaten der Meta-Analysen vor. Eine Ursache für diese Abweichungen können Rundungsdifferenzen bei der Nachberechnung der Effektstärkeschätzer und ihrer Präzisionsmaße aus den in den Meta-Analysen berichteten Primärstudiendaten sein. Durch eine unmittelbare Überprüfung eines Publikationsbias in jeder Meta-Analyse, wie in MARS und Prisma empfohlen, würde auch dieses Problem gelöst.

Die in der vorliegenden Arbeit verwendeten Methoden zur Überprüfung eines Publikationsbias verfügen weiterhin über eine geringe Teststärke, wenn nur wenige Primärstudien in einer Meta-Analyse eingeschlossen sind. Begg und Mazumdars Rangkorrelationstest (1994) und Eggers Regressionsanalyse (Egger et al., 1997) weisen bei weniger als zehn Primärstudien je Datensatz eine niedrige Teststärke auf (Sterne & Egger, 2005). Eine gute Teststärke liegt für die Trim und Fill Methode (Taylor & Tweedie, 2000a, 2000b) erst dann vor, wenn mehr als sechs Studien in einem Datensatz fehlen (Duval, 2005). In einigen der in dieser Arbeit untersuchten Datensätze waren weniger als zehn Studien

eingeschlossen, und es wurde zumeist nur eine geringe Anzahl fehlender Studien eingefügt. Deshalb ist die Teststärke für manche der vorliegenden Berechnungen als niedrig einzuschätzen, was zu dem geringen Ausmaß an Publikationsbias in den Ergebnissen beigetragen haben könnte.

Abschließend lässt sich festhalten, dass ein Publikationsbias in den untersuchten Meta-Analysen nur in geringem Maße vorliegt. Die Wirksamkeit der überprüften Interventionen ist mit Ausnahme einiger Therapiemethoden für die Schizophrenie substantiell gegeben. Die nicht von einem Publikationsbias betroffenen wirksamen Therapieverfahren sowie die nach der Korrektur um den Einfluss des Bias als wirksam ausgewiesenen therapeutischen Interventionen können in der klinischen Praxis zur Behandlung von Essstörungen, Schizophrenie und Depressionen, zur Reduktion der damit verbundenen Symptomatik und des psychischen Leidens, eingesetzt werden. Die Interventionen, deren Wirksamkeitseinschätzung sich durch die Korrektur um den Einfluss eines Publikationsbias verändert hat, wurden identifiziert und sollten unter Orientierung an den um den Einfluss eines Publikationsbias korrigierten Wirksamkeitsnachweisen in der klinischen Praxis angewendet werden.

Literatur

Meta-Analysen sind mit einem Asterisk (*) gekennzeichnet.

American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders (3rd ed.)* (DSM-III). Washington, DC: American Psychiatric Association

American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders (3rd ed. rev.).* Washington, DC: American Psychiatric Association.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders DSM-IV (4th ed.).* Washington, DC: American Psychiatric Association.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR (4th ed., rev.).* Washington: DC: American Psychiatric Association.

American Psychological Association (2009). *Publication manual of the American Psychological Association (6th ed.)*. Washington: DC American Psychiatric Association.

APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008). Reportings standards for research in psychology. Why do we need them? What might they be? *American Psychologist*, *43,* 839-851.

* Barbato, A. & D'Avanzo, B. (2008). Efficacy of Couple Therapy as a Treatment for Depression: A Meta-Analysis. *Psychiatric Quarterly, 79*, 121-132.

Begg, C.B. & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*, 1088-101.

*Beltman, M.W., Voshaar, R.C.O. & Speckens, A.E. (2010). Cognitive behavioural therapy for depression in people with a somatic disease: meta-analysis of randomized controlled trials. *The British Journal of Psychiatry, 197*, 11-19.

* Bola, J.R. (2006). Medication-free Research in Early Episode Schizophrenia: Evidence of Long-Term Harm? *Schizophrenia Bulletin, 32*, 288-296.

Borenstein M., Hedges L.V., Higgins J.P.T. & Rothstein H.R. (2005). *Comprehensive Meta-Analysis, Version 2.2.* Englewood, NJ: Biostat.

Borenstein M., Hedges L.V., Higgins J.P.T. & Rothstein H.R. (2009). *Introduction to meta-analysis*. West Sussex: Wiley & Sons.

* Bortolotti, B., Menchetti, M., Bellini, F., Montaguti, M.B. & Berardi, D. (2008). Psychological interventions for major depression in primary care: a meta-analytic review of randomized controlled trials. *General Hospital Psychiatry, 30*, 293-302.

Brownley, K.A., Berkman, N.D., Sedway, J.A., Lohr, K.N. & Bulik, C.M. (2007). Binge eating disorder treatment: A systematic review of randomized controlled trials. *International Journal of Eating Disorders*, *40*, 337-348.

Cooper, H., DeNeve, K. & Charlton, K. (1997). Finding the missing science: The fate of studies submitted for review by a human subjects committee. *Psychological Methods, 2,* 447-452.

Coursol, A. & Wagner, E.E. (1986). Effect of positive findings on submission and acceptance rates: A note on meta-analysis bias. *Professional Psychology Research and Practice, 17,* 136-137.

* Cuijpers, P., Brannmark, J.G. & Van Straten, A. (2008). Psychological treatment of postpartum depression: A meta-analysis. *Journal of Clinical Psychology, 64*, 103-118.

* Cuijpers, P., Munoz, R.F., Clarke, G.N. & Lewinsohn, P.M. (2009). Psychoeducational treatment and prevention of depression: the "coping with depression" course thirty years later. *Clinical Psychology Review, 29*, 449-458.

* Cuijpers, P., Smit, F., Bohlmeijer, E., Hollon, S.D. & Andersson, G. (2010). Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. *The British Journal of Psychiatry*, *196*, 173-178.

* Cuijpers, P., Van Straten, A. & Smit, F. (2006b). Psychological treatment of late-life depression: A meta-analysis of randomized controlled trials. *International Journal of Geriatric Psychiatry, 21*, 1139-1149.

* Cuijpers, P., Van Straten, A., Smits, N. & Smit, F. (2006a). Screening and early psychological intervention for depression in schools. Systematic review and meta-analysis. *European Child and Adolescent Psychiatry, 15*, 300-307.

* Cuijpers, P., Van Straten, A. & Warmerdam, L. (2007). Behavioral activation treatments of depression: A meta-analysis. *Clinical Psychology Review, 27*, 318-326.

* Cuijpers, P., Van Straten, A., Warmerdam, L. (2008). Are individual and group treatments equally effective in the treatment of depression in adults? A meta-analysis. *The European Journal of Psychiatry, 22*, 38-51.

Degkwitz, R., Helmchen, H., Kockott, G. &Mombour, W., Hrsg. (1980). *Diagnosenschlüssel und Glossar psychiatrischer Krankheiten: dt. Ausgabe d. internat. Klassifikation d.*

Krankheiten d. WHO, ICD, 9. Revision. Chapter V. Springer, Berlin, Heidelberg, New York.

* De Maat, S., Dekker, J., Schoevers, R. & De Jonghe, F. (2006). Relative efficacy of psychotherapy and pharmacotherapy in the treatment of depression: A meta-analysis. *Psychotherapy Research, 16*, 562-672.

* De Maat, S.M., Dekker, J., Schoevers, R.A. & De Jonghe, F. (2007). Relative efficacy of psychotherapy and combined therapy in the treatment of depression: A meta-analysis. *European Psychiatry, 22*, 1-8.

* Dennis, C.-L. (2005). Psychosocial and psychological interventions for prevention of postnatal depression: systematic review. *British Medical Journal, 331*, 1-8.

Dickersin, K. (1997). How important is publication bias? A synthesis of available data. *AIDS Education and Prevention, 9 (1 Suppl.),* 15-21.

Dickersin, K. (2005). Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis,* pp. 11-34. West Sussex: Wiley & Sons Ltd.

Dickersin, K., Chan, S., Chalmers, T.C., Sacks, H.S. & Smith, H. Jr. (1987). Publication bias and clinical trials. *Controlled Clinical Trials, 8*, 343-353.

Duval, S. (2005). The Trim and Fill method. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 127-144). West Sussex: Wiley & Sons Ltd.

* Ekers, Richards & Gilbody (2008). A meta-analysis of randomized trials of behavioural treatment of depression. *Psychological Medicine, 38*, 611-623.

Egger M., Davey Smith, G., Schneider, M. & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629-34.

Epstein, W.M. (1990). Confirmational response bias among social work journals. *Science, Technology, and Human Values, 15,* 9-37.

Ernst, E. & Pittler, M.H. (1997). Alternative therapy bias [letter]. *Nature, 385*, 480.

Fydrich, T. & Schneider, W. (2007). Evidenzbasierte Psychotherapie. *Psychotherapeut, 52*, 55-68.

Glass, G.V., McGaw, B. & Smith, M.L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage Publications.

Greenwald, A.G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82*, 1-20.

* Haby, M.M., Donnelly, M., Corry, J. & Vos, T. (2006). Cognitive behavioural therapy for depression, panic disorder and generalized anxiety disorder: A meta-regression of factors that may predict outcome. *Australian and New Zealand Journal of Psychiatry, 40*, 9-19.

Hahlweg, K. & Dose, M. (1998). *Schizophrenie. Fortschritte der Psychotherapie, Bd. 2.* Göttingen: Hogrefe Verlag.

Hautzinger, M. (1998). *Depression. Fortschritte der Psychotherapie, Bd. 4*. Göttingen: Hogrefe Verlag.

* Hay, P.P.J., Bacaltchuk, J., Stefano, S. & Kashyap, P. (2009). Psychological treatments for bulimia nervosa and binging (Review). *The Cochrane Library*, 4.

Herpertz, S., Herpertz-Dahlmann, B., Fichter, M., Tuschen-Caffier, B. & Zeeck, A. (2011). *S3-Leitlinie Diagnostik und Behandlung der Essstörungen.* Heidelberg: Springer Verlag.

* Himelhoch, S., Medoff, D. & Oyeniyi, G. (2007). Efficacy of group psychotherapy to reduce depressive symptoms among HIV-infected individuals: A systematic review and meta- analysis. *Aids Patient Care and STDs, 21*, 732-739.

Hoek, H.W. (2006). Incidence, prevalence and mortality of anorexia nervosa and other eating disorders. *Current Opinion in Psychiatry*, *19*, 389-394.

Hopewell, S., Clarke, M. & Mallett, S. (2005). Grey literature and systematic reviews. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis,* pp. 49-72. West Sussex: Wiley & Sons Ltd.

Hunter, J.E. & Schmidt, F.L. (2000). Fixed effects vs. Random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8,* 275-292.

Ioannidis, J.P.A. & Trikalinos, T.A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *Canadian Medical Association Journal*, *176,* 1091-1096.

Jacobi, C., Paul, T. & Thiel, A. (2004). *Essstörungen. Fortschritte der Psychotherapie, Bd. 24*. Göttingen, Hogrefe.

Jennions, M.D. & Møller, A.P. (2002). Publication bias in ecology and evolution: An empirical assessment using the 'trim and fill' method. *Biological Review, 77,* 211-222.

* Jones, C., Cormac, I., Silveira da Mota Neto, J. I. & Campbell, C. (2010). Cognitive behaviour therapy for schizophrenia. *The Cochrane Database of Systematic Reviews, 4*, 1-69.

Klerman, G.L., Weissman, M.M., Rounsaville, B.J. & Chevron, E.S. (1984). *Interpersonal Psychotherapy of Depression*. New York: Basic Books.

Legenbauer, T. & Vocks, S. (2006). *Manual der kognitiven Verhaltenstherapie bei Anorexie und Bulimie.* Heidelberg: Springer Medizin Verlag.

Light, R.J. & Pillemer, D.B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

* Lincoln, T.M., Suttner, C. & Nestoriuc, Y. (2008). Wirksamkeit kognitiver Interventionen für Schizophrenie. Eine Meta-Analyse. *Psychologische Rundschau, 59*, 217- 232.

Lipsey, M.W. & Wilson, D.B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181-1209.

* Lynch, D., Laws, K.R. & McKenna, P.J. (2010). Cognitive behavioural therapy for major psychiatric disorder: Does it really work? A meta-analytical review of well-controlled trials. *Psychological Medicine, 40*, 9-24.

Mahoney, M.J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy & Research, 1*, 161-175.

McGrath, J., Saha, S., Chant, D. & Welham, J. (2008). Schizophrenia: A concise overview of incidence, prevalence, and mortality. *Epidemiologic Reviews, 30*, 67-76.

Moher, D., Liberati, A., Tetzlaff, J. & Altman, D.G. (2009). Preferred reporting items for systematic reviews and meta-analysis: The PRISMA statement. *Annals of Internal Medicine*, *151*, 264-270.

Moscati, R., Jehle, D., Ellis, D., Fiorello, A. & Landi, M. (1994). Positive-outcome bias: Comparison of emergency medicine and general medicine literatures. *Academic Emergency Medicine, 1*, 267-271.

Mulrow, C.D. (1994). Rationale for systematic reviews. *British Medical Journals, 309*, 597-599.

* Neumeyer-Gromen, A., Lampert, T., Stark, K. & Kallischnigg, G. (2004). Disease management programs for depression: A systematic review and meta-analysis of randomized controlled trials. *Medical care, 42,* 1211-21.

* Newton-Howes, G., Tyrer, P. & Johnson, T. (2006). Personality disorder and the outcome of depression: Meta-analysis of published studies. *British Journal of Psychiatry, 188*, 13-20.

* Pampallona, S., Bollini, P., Tibaldi, G., Kupelnick, B. & Munizza, C. (2004). Combined pharmacotherapy and psychological treatment for depression. *Arch Gen Psychiatry, 61*, 714-719.

* Pharoah, F., Mari, J., Rathbone, J. & Wong, W. (2010). Family intervention for schizophrenia (Review). *The Cochrane Library, 12*, 1-156.

* Pitschel-Walz, G., Leucht, S., Bäuml, J., Kissling, W. & Engel, R. (2001). The effect of family interventions on relapse and rehospitalization in schizophrenia - A meta-analysis. *Schizophrenia Bulletin, 27*, 73-92.

* Pitschel-Walz, G. (1997). Die Einbeziehung von Angehörigen in die Behandlung schizophrener Patienten und ihr Einfluss auf den Krankheitsverlauf. *Europäische Hochschulschriften, Reihe 6, Vol. 581.* Frankfurt a.M.: Peter Lang Verlag.

* Pitschel-Walz, G. & Engel, R. (1997). Psychoedukation in der Schizophreniebehandlung; Die Einbeziehung der Angehörigen in die Behandlung schizophrener Patienten und ihr Einfluss auf den Krankheitsverlauf. *Europäische Hochschulschriften, 23.* Frankfurt a.M.: Peter Lang Verlag.

Pulverich, G. (1999). *Psychotherapeutengesetz*. Kommentar, 3. Aufl. Bonn: Deutscher Psychologen Verlag.

* Reinecke, M.A., Ryan, N.E. & DuBois, D.L. (1998). Cognitive-behavioral therapy of depression and depressive symptoms during adolescence: A review and meta-analysis. *J.Am.Acad.Child Adolesc. Psychiatry, 37*, 26-34.

Roder, V., Brenner, H.D., Kienzle, N. & Hodel, B. (1997). *Integriertes psychologische Therapieprogramm für schizophrene Patienten (IPT).* München, Weinheim: Psychologie Verlags Union.

Rothstein, H.R., Sutton, A.J. & Borenstein, M. (2005). Publication bias in meta-analysis. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 1-8). West Sussex: Wiley & Sons Ltd.

Rustenbach, S.J. (2003). *Metaanalyse. Eine anwendungsorientierte Einführung*. Verlag Hans Huber, Bern.

Scherer, R.W., Langenberg, P. & Van Elm, E. (2005). The Cochrane Library (2005). Full Publication of results initially presented in abstracts. *The Cochrane Database of Methodology Reviews 2005, Issue 2*. Art No.: MR000005. DOI: 10.1002/14651858.MR000005.pub2. John Wiley & Sons, Ltd.

Shadish, W.R., Doherty, M. & Montgomery, L.M. (1989). How many studies are in the file drawer? An estimate from the family/marital psychotherapy literature. *Clinical Psychology Review, 9*, 589-603.

Singleton, N., Bumpstead, R., O'Brien, M., Lee, A. & Meltzer, H. (2001). *Psychiatric morbidity among adults living in private households (2000).* London: The Stationery Office.

Slade, M. & Priebe, S. (2001). Are randomized controlled trials the only gold that glitters? BJP 2001, 179, 286-287.

Smith, M.L. (1980a). Publication bias and meta-analysis. *Evaluating Education, 4,* 22-24.

Smith, M.L. (1980b). Sex bias in counselling and psychotherapy. *Psychological Bulletin, 87,* 392-407.

Song, F., Eastwood, A., Gilbody, S., Duley, L. & Sutton, A. (2000). Publication and other selection biases in systematic reviews. *Health Technology Assessment, 4*, 1-115.

Song, F., Khan, K.S., Dinnes, J. & Sutton, A. (2002). Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *International Journal of Epidemiology, 31,* 88-95.

Sterling, T.D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance- or vice versa. *Journal of the American Statistical Association, 54*, 30-34.

Sterne, J.A.C., Becker, B.J. & Egger, M. (2005). The funnel plot. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 75-98). West Sussex: Wiley & Sons Ltd.

Sterne, J.A.C. & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 99-110). West Sussex: Wiley & Sons Ltd.

Sterne, J.A.C., Gavaghan, D. & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology, 53*, 1119-1129.

Sterne, J.A.C., Jüni, P., Schulz, K.F., Altman, D.G., Bartlett, C. & Egger, M. (2002). Statistical methods for assessing the influence of study characteristics on treatment effects in ´meta-apidemiological´ research. *Statistics in Medicine, 21*, 1513-1524.

*Stice, E. & Shaw, H. (2004). Eating disorder prevention programs: A meta-analytic review. *Psychological Bulletin*, *130*, 206-227.

Sutton, A.J. (2005). Evidence concerning the consequences of publication and related biases. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), Publication bias in meta-analysis. West Sussex: Wiley & Sons Ltd, pp. 175-192.

Sutton, A.J., Duval, S.J., Tweedie, R.L., Abrams, K.R. & Jones, D.R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *British Medical Journal, 320*, 1574-1577.

Taylor, S. & Tweedie, R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95,* 89-98.

Taylor, S. & Tweedie, R. (2000b). Trim and Fill: A simple funnel-plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455-63.

Terrin N., Schmid, C.H., Lau, J. & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine*, *22*, 2113–2126.

Vickers, A., Gloyal, N., Harland, R. & Rees, R. (1998). Do certain countries produce only positive results? A systematic review of controlled trials. *Controlled Clinical Trials, 19*, 159-166.

Vocks, S., Tuschen-Caffier, B., Pietrowsky, R., Rustenbach, S. J., Kersting, A. & Herpertz, S. (2010). Meta-analysis of the effectiveness of psychological and pharmacological treatments for binge eating disorder. *International Journal of Eating Disorders*, *43*. 205-217.

White, K.R. (1982). The relationship between socioeconomic status and academic achievement. *Psychological Bulletin, 91*, 461-481.

Wolf, F.M. (1990). Methodological Observations on Bias. In: Wachter, K.W. & Straf, M.L. (Eds.), *The Future of Meta-Analysis*, pp. 139-151. New York, NY: Russell Sage.

World Health Organisation (1992). *The ICD-10 classification of mental and behavioral disorders: Clinical descriptions and diagnostic guidelines*. Geneva: Author.

* Wykes, T., Steel, C., Everitt, B. & Tarrier, N. (2008). Cognitive behavior therapy for schizophrenia: Effect sizes, clinical models, and methodological rigor. *Schizophrenia Bulletin, 34*, 523-537.

* Zimmermann, G., Favrod, J., Trieu, V.H. & Pomini, V. (2005). The effect of cognitive behavioral treatment on the positive symptoms of schizophrenia spectrum disorders: A meta- analysis. *Schizophrenia Research, 77*, 1-9.

Einzelarbeiten

Untersuchung 1

Niemeyer, H., Bongartz, L., Musch, J. & Pietrowsky, R. (2011). Publication bias in meta-analyses of the efficacy of psychotherapeutic interventions for eating disorders (manuscript under review).

Untersuchung 2

Niemeyer, H., Musch, J. & Pietrowsky, R. (2011). Publication bias in meta-analyses of the efficacy of psychotherapeutic interventions for schizophrenia (manuscript under review).

Untersuchung 3

Niemeyer, H., Musch, J. & Pietrowsky, R. (2011). Publication bias in meta-analyses of the efficacy of psychotherapeutic interventions for depression (manuscript under review).

Submission Confirmation

Thank you for submitting your manuscript to *International Journal of Eating Disorders*.

Manuscript ID:	IJED-11-0257
Title:	Publication Bias in Meta-Analyses of the Efficacy of Psychotherapeutic Interventions for Eating Disorders
Authors:	Niemeyer, Helen Bongartz, Lisa Musch, Jochen Reinhard, Pietrowsky
Date Submitted:	13-Aug-2011

Helen Niemeyer, MSc, Heinrich-Heine-Universität Düsseldorf

Lisa Bongartz, Heinrich-Heine-Universität Düsseldorf

Jochen Musch, PhD, Heinrich-Heine-Universität Düsseldorf

Reinhard Pietrowsky, PhD, Heinrich-Heine-Universität Düsseldorf

PUBLICATION BIAS IN META-ANALYSES

Abstract

Objective: Publication bias should be assessed in every meta-analysis to ensure the validity of the results. The aim of this study was to assess whether systematic reviews investigating psychotherapeutic interventions for eating disorders are affected by publication bias.

Method: All systematic reviews up to September 2010 which reported the necessary data to assess publication bias were included. Begg and Mazumdar´s adjusted rank correlation test, Egger´s regression analysis and the Trim and Fill procedure were applied.

Results: We analysed 14 data sets reported in 2 meta-analyses for hints of the existence of publication bias. Begg's test indicated significant bias in two (14.29 %) of these data sets, Egger's test in three (21.43%). The Trim and Fill procedure changed the evaluation of the efficacy of an intervention only once (7.14%).

Discussion: We found little evidence for the presence of publication bias. With one notable exception, correcting for publication bias did not change the evaluation of the efficacy of psychotherapeutic interventions.

Publication Bias in Meta-Analyses of the Efficacy of Psychotherapeutic Interventions for Eating Disorders

Eating disorders are characterized by severe psychic symptoms, prevalence rates in the general population are 0.3-1% for bulimia nervosa (BN), 0.3-3% for anorexia nervosa (AN) (1; 2), and 0.7-6.6 % for binge eating disorder (BED; 3; 4). The specific psychopathology underlying eating disorders comprises concerns about weight and shape and also severely affects self-esteem (5). The core symptoms of AN are starvation and weight loss. While binge eating is the characteristic symptom of BED, compensatory behavior such as purging is the defining additional characteristic of BN. Dietary restraints and body-image disturbances are characteristic for BN and AN, whereas obesity is common in BED. High comorbidity rates, especially for depression, are characteristic of all eating disorders (6; 7). For clinically significant eating disorder symptoms which do not fulfill all criteria for one of the above disorders, the additional category “eating disorders not otherwise specified“ (EDNOS) has been established in the DSM-IV (8). Furthermore, almost 10% of young women suffer from subthreshold and threshold eating disorder symptoms (9) and are thus at a high risk of developing a full disorder.

Untreated eating pathology often becomes chronic, and even after treatment the course of the disorder is chronic or intermittent in 25 % of cases for AN (10), 40 % for BN (11) and 43% for BED (12). Efficacious treatments are essential to alleviate both the distress and the functional impairment caused by eating disorders, and to reduce both the high prevalence and chronicity rates.

Cognitive behavioral therapy (CBT) and interpersonal therapy (IP) are the two most frequently investigated therapeutic interventions for eating disorders. Further therapies include dialectic behavioral therapy (DBT), psychoanalysis, psychodynamic therapy, humanistic therapy, systemic therapy and family therapy. In addition, in order to avert the

development of threshold eating symptoms into full eating disorders, prevention programs for universal as well as high risk samples have been developed. Therapeutic interventions and prevention programs differ with regard to their theoretical conceptualization of eating disorders, their therapeutic approaches and mechanisms, and their approach to forming a patient-therapist relationship. Consequently, there are differences in their efficacy. In order to reduce the suffering resulting from eating disorders, the most effective techniques need to be offered to patients, as well as to people with onset symptoms to prevent the development of a full disorder.

A large number of studies have been conducted to assess the efficacy of psychotherapeutic interventions to treat eating disorders. Because it is difficult for researchers and practitioners to obtain an overview over this multitude of studies, which sometimes also include contradictory results, meta-analyses have been used for about two decades to synthesize the results (13). Because meta-analyses are based upon larger samples, they provide comprehensive and more precise evidence and thus have increased inferential power (14).

The integration of empirical evidence in meta-analyses relies mainly on published studies which are considerably easier to retrieve and, having passed a peer review process, are often considered to represent higher quality research than unpublished studies. This rationale is hardly convincing, however. As Lipsey and Wilson (15) point out, unpublished studies may be of the same quality, and merely integrating published results may lead to systematically biased effect size estimates that are not representative of the entire body of research that has been conducted in a particular domain. All primary studies, including unpublished ones, should therefore be included in a meta-analysis in order to obtain valid and unbiased results. Otherwise, the set of studies that is included for meta-analysis may suffer from significant bias (14).

Sterling (16) appears to have been the first to suggest that positive results are more likely to be published. But if the magnitude and the direction of an effect influence the decision to submit or accept a research report, this induces a positivity bias into the published body of research. Smith (17) coined the term *publication bias* for the selective publication of positive or statistically significant results (18). There is consensus among most researchers that publication bias has to be considered one of the biggest threats to the validity of meta-analyses (19; 14).

In a literature search outlined in more detail below, we found that 17 meta-analyses have been conducted to assess the efficacy of psychotherapeutic interventions for eating disorders including BN, BED and EDNOS, and the prevention of eating disorders up to September 2010 (4; 13; 20-34). The therapeutic approaches that were assessed in these meta-analyses are CBT, exposure and response prevention, IP, DBT, psychodynamic therapy, analytic therapy, insight oriented therapy, non-directive therapy, nutritional and weight management, psychoeducation, stress management and dissonance induction techniques. In two systematic reviews of AN, no integrated mean effect size was reported (35; 36). None of the 17 meta-analyses conducted to assess the efficacy of therapeutic interventions for eating disorders controlled for publication bias with methods such as Trim and Fill (37; 38), Begg and Mazumdar´s rank correlation test (39) or Egger´s regression analysis (40), indicating a wide-spread lack of statistical control for publication bias. Furthermore, only 6 of the 17 meta-analyses (35 %) included unpublished studies. The importance of including unpublished studies or assessing publication bias was not discussed in the other eleven meta-analyses. While including only published studies may be defensible to the extent that the quality of published studies is indeed better than that of unpublished studies, limiting a meta-analysis to published studies may well introduce a systematic bias and can therefore seriously threaten the validity of the meta-analysis (14). Consequently, undetected publication bias may exist in

the meta-analyses that have been published to assess the efficacy of therapeutic inventions, and the efficacy of the interventions mentioned above may therefore have been overestimated.

An overestimation of the efficacy of therapeutic interventions is especially problematic if biased empirical evidence convinces researchers and psychotherapists to endorse interventions that are mistakenly supported in meta-analyses (39). As a consequence, publication bias is not just a statistical problem, but may also have serious practical implications. The possible negative consequences are probably most severe for clinicians and patients: Therapeutic techniques that are considered to be most effective and helpful are taught to psychotherapists, and, as a consequence, patients with eating disorders might be treated with inefficient or even harmful techniques as a result of biased meta-analyses. This may also lead to unnecessarily high costs, not only for the health care system, but also for society at large. Publication bias may also mislead researchers by inducing them to conduct studies dealing with questions that have already been investigated, but were not published because of their non-significant result. This may cause a waste of money, time and personnel resources, and must also be considered unethical with regard to the patients participating in such studies.

The selective publication of studies with positive results may occur for a number of different reasons, including the limited space available in scientific journals, explicit or implicit editorial policies to regard significant results as more meaningful, and the reluctance of researchers to invest work in submitting articles reporting negative or null findings in anticipation of their likely rejection. Other reasons for investigators to selectively publish significant results might be that the results of a study favor a researchers' or publishers' preferred intervention, or that they consider them to be more “interesting” or "convincing". Finally, it is also possible that ongoing studies are stopped if it becomes foreseeable that their results will be insignificant (18).

Selective reporting can also occur as a result of small study effects. Minor effects do not reach significance if they are investigated in small samples, even though an effect of similar size would become significant in a larger sample. If the significance of a study determines the probability of its publication, small studies are more likely to get published if they yield larger effects. This may result in a correlation between study size and effect size (41). Accordingly, Kraemer and colleagues (42) argued that publication bias can occur as a result of the selective publication of significant and extreme effects which are based on small sample sizes with insufficient power. This poses a problem to the meta-analytic integration of research findings because small studies are more likely to show large effects by chance (14).

The perhaps best solution to prevent publication bias is to implement registers in which all studies have to be listed at the time of their inception (43; 14). Another approach is that researchers conducting meta-analyses take pains to also include unpublished studies. However, since unpublished studies are difficult to obtain, this is usually not feasible. As a consequence, publication bias should be controlled for statistically in every meta-analysis (14; 44).

Meta-analyses of psychotherapeutic interventions for eating disorders have not been comprehensively assessed for publication bias. Publication bias might therefore be an as yet undiscovered threat to the validity of the empirical evidence regarding the efficacy of interventions for eating disorders. The aim of this study was to assess the influence of publication bias in meta-analyses investigating psychotherapeutic interventions for eating disorders. The existence and impact of a possible publication bias was tested, and in case bias was found, corrected effect size estimates were calculated. In addition, the difference between the original and the unbiased effect sizes was tested for significance.

METHODS

Data sources

In order to assure the comprehensive retrieval of all published meta-analyses and systematic reviews about psychotherapeutic interventions for eating disorders, we followed the search strategies recommended by Lipsey & Wilson (15). First, we performed an extensive computer search using the databases PsycInfo and Psyndex up till September 2010. The keywords “meta-analysis” or “systematic review” and ”eating disorders” were used in combination, as well as combinations of “meta-analysis” or “systematic review” and each of the main diagnoses, “bulimia nervosa”, “anorexia nervosa” and “binge eating disorder”. The search was confined to reviews and meta-analyses reported in English or German language. Cross-references to other articles and book chapters were also followed if they were found in the initial search.

Study selection and data extraction

When an abstract provided insufficient information, the respective article was checked in order to not miss a relevant review. Eligible systematic reviews and meta-analyses had to involve BN, AN, BED, or EDNOS in general according to the diagnostic criteria of either DSM-III-R (45), DSM-IV (8), DSM-IV-TR (46), ICD-9 (47) or ICD-10 (48). Meta-analyses and reviews regarding prevention of eating disorders with unselective or high risk samples also met the inclusion criteria. Eligible articles had to involve at least one therapeutic intervention. All psychotherapeutic interventions employed were included. Reviews that targeted several disorders simultaneously and were not restricted to the treatment of eating disorders were excluded. Finally, we excluded meta-analyses that examined the effectiveness of pharmacological treatment because the focus of this study is on psychotherapy research.

Some meta-analyses reported several arms of studies. In this case, we restricted our analyses to those arms reporting a pooled effect size estimator, the primary studies´ effect

sizes and a measure of their precision, because these data are necessary to assess publication bias. If these data were not reported, we attempted to obtain them from the authors. A study was included if some measure of precision (confidence intervals, standard deviations, standard errors or variances) was available for each study. If the effect size index was a correlation coefficient, no measure of precision was needed if the sample size was reported instead. If any of the necessary data was missing, the raw data means and standard deviations or variances – if available - were used to calculate the effect sizes and their precision. Cases in which the reported effect sizes did not match the effect sizes we computed on the basis of the raw data were excluded from analysis.

When using methods based on an investigation of funnel-plot asymmetry to detect the presence of publication bias, it is important to note that the cause of asymmetry cannot by determined from the asymmetry itself. Asymmetry may result from many causes, including heterogeneity, undetected covariates, methodological inadequacies, or chance (49). In case of between-study heterogeneity, statistical methods based on a homogeneity assumption are inappropriate and lead to false alarms (37, 38, 50, 51, 52). Therefore, to avoid erroneous conclusions due to a violation of the assumptions underlying the statistical tests to detect publication bias, homogeneity of the pooled effect sizes was set as an additional inclusion criterion. Finally, since an assessment of publication bias is unreliable and does not have sufficient power if the number of studies is too small (41), all study arms comprised of less than six studies were also excluded.

Assessment of publication bias

Several methods for the assessment of publication bias have been developed. The most familiar is the funnel plot (53), a scatter plot with the effect size of the studies graphed on the abscissa against a measure of the size of the study or the precision of the effect plotted on the ordinate. If no publication bias is present, the effect sizes should be dispersed symmetrically

around the mean effect size estimate at any level of precision, and thus resemble an inverted funnel. The smaller and less precise studies are expected to scatter more widely at the bottom of the graph due to their poor precision, whereas the larger studies are expected to show less scatter at the top of the graph. If unpublished studies with null or negative effects are not included in the meta-analysis, the dispersion becomes asymmetrical because of a gap in the funnel where the smaller studies with null or negative effect sizes would have to be expected but are missing. The assessment of publication bias based on a visual analysis of the funnel plot simply takes a look at the asymmetry in the dispersion of the effect sizes around their mean. The major drawback of this method is the subjectivity of a merely visual interpretation of asymmetry. The visual analysis is also confounded by the fact that different measures of precision lead to very different shapes of the funnel plot (41). To resolve this problem, quantitative and thus more objective methods have been developed to detect publication bias or small study effects. The three methods that were applied in the present study are described below.

The first method we applied is a non-parametric test to detect small study effects, the Begg and Mazumdar (39) adjusted rank correlation method. It examines the association between the standardized effect sizes and their variances using Kendall´s Tau. Ranks are assigned separately for both measures, and a trend toward publication of studies with smaller samples and large effects is indicated by a significant positive correlation between the effect sizes and their variances. If there is no publication bias, no relationship between the effect sizes and their variance is to be expected. Testing the correlation for significance therefore provides a statistical test for small study effects and hence, publication bias.

The second method is a parametric test with higher power to detect small study effects, Egger´s linear regression method (40). The standard normal deviate (the effect size standardized by its standard error) is regressed against its precision (the inverse of its standard

error). The slope represents the magnitude and direction of the effect, and the intercept yields a measure of asymmetry in a funnel plot: if the regression line does not run through the origin, this indicates asymmetry in the dispersion and the intercept can be used as a measure of the degree of asymmetry. Both Begg´s and Egger´s methods do not adjust for publication bias. Two-sided p-values should be utilized, as Sterne and Egger (54) argue, because a reverse pattern with larger studies showing larger effects might also occur. However, other authors have argued that one-tailed tests should be conducted, because publication bias is expected to reduce, rather than to enlarge the mean effect size (55). We therefore decided to report both one- and two-sided significance tests. The probability of a type I error was set at.05.

The third method, the non-parametric Trim and Fill procedure (37; 38), goes beyond the mere assessment of the degree of asymmetry in a funnel plot; rather, it estimates and adjusts for the number of apparently missing studies, as well as their potential impact on the estimate of mean effect size. In the absence of bias the effect sizes are distributed symmetrically around the true effect size. In case of asymmetry, the probable number of missing studies is estimated by first trimming off the asymmetric right-hand side of the funnel plot, which consists of the largest effect sizes, and second re-filling the trimmed studies and the imputed missing studies with null and negative effects on the left-hand side around the symmetric remainder. By means of this procedure, Trim and Fill estimates the number and outcome of missing studies and a new mean effect size, adjusted for bias. The trimming and filling is done by an algorithm that usually stabilizes after a few iterations, when the resulting funnel plot is symmetric. The corrected effect size estimate is based on the data set of the observed and the augmented studies, and a new 95% confidence interval based on the observed and the imputed studies is computed. Trim and Fill should be seen as a sensitivity analysis for the robustness of the effect estimate against publication bias, and one should not

interpret the number and effect sizes of the missing studies generated by Trim and Fill as the exact values of the missing studies (56).

All three methods suffer from reduced power to detect small-study effects if they are based on a small number of studies (37; 38; 41; 54; 56). The three methods give comparable results for the detection of bias in most cases (56; 57), but they might differ in some cases due to their different power and an unequal risk for false positives, depending on the number of primary studies the analysis is based on and their sample size.

All tests were performed using the software "Comprehensive Meta-Analysis" (Version 2.2), which was developed for meta-analytic effect calculations (58). To apply Begg´s rank correlation method, Egger´s regression analysis and the Trim and Fill procedure, the primary studies´ effect sizes and a measure of their precision were entered and the mean effect size was calculated in an intermediate step.

Two additional methods that have been proposed to detect publication bias, the failsafe N (59; 60) and an approach based on selection models (61), were not applied in the present study. This decision was made in the case of the failsafe N formula because of its many shortcomings. Firstly, the assumption underlying Rosenthal's formula (59) is that the average effect size of missing results is null, which is quite unlikely if they are missing due to publication bias (62). Furthermore, the formula is just an „ad hoc rule of thumb" (60), as neither the sample size of the missing studies (studies with n=10 are set equal to studies with n=1000) nor information about a possible heterogeneity or potential moderators or other study characteristics can be accounted for. No statistical model underlies the formula, and the distribution of the failsafe N statistic is unknown. Failsafe N and related formulas have been developed by a number of authors (63-65), but none of them solves all of the problems outlined above. Moreover, different formulas for the computation of the failsafe N lead to

different estimates of the number of missing studies. Becker (60) therefore concluded that the method should be abandoned in favor of more informative methods (57).

In the case of selection models, the probability of publication is modeled as a weighted function of the p-values associated with each primary study, or as a function of the primary studies´ effect size estimates and their associated standard errors. The number of missing studies and their effect on the pooled effect size are estimated. The major disadvantage of selection models is that they are computationally burdensome and highly computer intensive to run. For these reasons they are not frequently used (61; 66) and are not applied here.

RESULTS

Description of meta-analyses investigated

Our search yielded two meta-analyses that met all inclusion criteria and thus qualified for the present analysis: those of Hay, Bacaltchuk, Stefano and Kashyap (23), and Stice & Shaw (28).

The efficacy of cognitive behavioral therapy (CBT), interpersonal therapy, hypno-behavioral therapy, dialectic behavioral therapy, supportive therapy and any other psychotherapy including behavioral weight loss treatment for overweight binge eaters was assessed in the meta-analysis of Hay et al. (23). The specific disorders under investigation were BN, BED and EDNOS. Individual effect sizes were calculated separately for two categories: for CBT on the one hand, and on the other hand for a category consisting of all other treatments. Only randomized controlled trials were included by Hay et al. (23). The relative risk of binge eating at the end of treatment and measures of the severity of bulimic symptoms were used as main outcomes, and additional secondary outcomes such as weight, depression, eating attitude, remission and dropout rates, interpersonal functioning and general psychiatric symptoms were also provided.

In the second meta-analysis by Stice and Shaw (28), prevention programs for eating disorders were analyzed. The risk status of the samples in the primary studies was given and classified as universal if the intervention was delivered to classes, or if the objective was not known to participants at the time of inception. Samples were classified as high-risk when participants were screened for risk factors and thus constituted a preselected sample. Both interactive and didactic formats of prevention programs were investigated. Overall effect sizes were also calculated by including all samples and formats. Only randomized controlled trials were included. Outcome measures consisted of eating pathology and five risk factors for eating disorders, namely thin-ideal internalization, body dissatisfaction, dieting, negative affect and body mass.

Both meta-analyses included unpublished studies. Publication bias had not been assessed statistically in either study, but Hay et al. (23) used funnel plots to detect asymmetry. Their visual examination of the funnel plots suggested some asymmetries which were briefly discussed in the original analysis, but no funnel plots were displayed for the separate outcomes, and no statistical test of the extent of the asymmetries was conducted.

The treatment efficacy was estimated by the standardized mean difference (*SMD*) and the relative risk (*RR*) in Hay et al. (23), who reported 95 % confidence interval limits as a measure of precision. Relative risks smaller than 1 indicated that a treatment was efficacious in the experimental group (49; 23). Stice and Shaw (28) provided correlation coefficients (*r*) and the size of the samples for which they were computed. Both Hay et al. (23) and Stice and Shaw (28) used the random effects model for the integration of study effects. We therefore also used the random effects model. We separately analyzed all mean effect sizes for data sets that were based on the integration of six or more primary studies´ effect sizes (41). Overall, 14 pooled effect sizes met the inclusion criteria of a) assessing the efficacy of therapeutic interventions or prevention programs for eating disorders, b) providing the effect sizes of the

included primary studies and a measure of their precision, c) homogeneity of the effects, and d) a data base of at least six studies (see Table 1). Eleven of these 14 data sets were originally reported by Hay et al. (23); three were reported by Stice and Shaw (28). A complete list of all 14 data sets is given in Table 1, which also reports the number of studies included in each data set. All of the eleven data sets reported by Hay et al. (23) refer to post treatment measures for different therapies and control groups; for example, one set of data compared patients treated with CBT and a waiting list control group regarding the severity of bulimic symptoms as the outcome variable. Each of the eleven sets of effect sizes included data from between one and four unpublished primary studies.

The three sets of effect sizes reported by Stice and Shaw (28) refer to follow-up effects. Two of the data sets also included 3 and 4 unpublished studies, respectively. To avoid false positives, we decided not to include the post-treatment effects also reported by Stice and Shaw (28), because we found these effect sizes to be heterogeneous using a Q-test, offering an obvious alternative explanation for the apparent asymmetry of the respective effect sizes.

For most of the mean effect sizes of the two meta-analyses, our recalculations performed on the raw data matched the original ones. In a few cases, however, the recalculations showed some – always minor - deviations from the original estimates (by .01 for one of the mean differences d, by .01 for one relative risk, and by .01, .03 and .04 for three correlation coefficients (*r*)). These small discrepancies were probably the result of different rounding. For a statistical assessment of publication bias, we used the recalculated effect sizes. The original and recalculated effect sizes are shown in Table 1.

Results of the Trim and Fill calculations

According to the Trim and Fill calculations, at least one missing study was identified in one or more of the data sets reported by Hay et al. (23) and by Stice and Shaw (28).

Corrected mean effect size estimates were calculated for all asymmetric data sets, taking the values of the putative missing cases into account. Overall, 9 of the 14 effect sizes were slightly altered by this procedure, because Trim and Fill identified up to four missing studies, thus indicating the presence of publication bias. In 4 data sets the therapeutic approaches were assessed as significant both before and after correction, and in 4 data sets the mean effect sizes were non-significant before and after reduction. However, in one data set the mean effect size changed from non-significance to significance after correcting for publication bias. The detailed results of the Trim and Fill calculations for all outcome measures are shown in Table 1.

Insert Table 1 about here

Data sets #1 to #3 and #5 to #10 in Table 1 estimated the efficacy of CBT for BN, BED and EDNOS combined (23). This was done either in comparison to a group of other psychotherapies taken together (PT), or as compared to the outcomes observed in a waiting list (WL) control group, using either *bulimic symptoms, weight, general psychiatric symptoms, interpersonal functioning, remission rates* or *dropout rates* as outcome measure. Data sets #4 and #11 estimated the efficacy of PT as compared to WL for the treatment of BN, BED and EDNOS using either *bulimic symptoms* or *dropout rates* as outcome measure. In six of the eleven data sets reported by Hay et al. (23), mean effect size estimates were based on mean standardized differences (*d*), three of which showed evidence of publication bias because additional studies were imputed by the Trim and Fill procedure. This included a set of studies concerned with the difference between CBT and PT regarding the reduction of *bulimic symptoms* (data set #1) and *overweight* (data set #2), and a set of studies investigating the efficacy of CBT as compared to a waiting list using *bulimic symptoms* (data set #3) as the outcome of interest. However, in spite of the additional studies imputed by the Trim and Fill procedure, there was no significant difference between the original effect size estimates and

the corrected ones in any of these three data sets. Moreover, contrary to expectations, only one of the imputed sets of studies (those for a comparison between CBT and PT regarding post-treatment *overweight*, data set #2) led to a reduced, rather than enlarged effect size estimate. Both before and after correction, the confidence interval for the effect size estimate for the comparison of CBT and PT included zero and thus, the correction according to the Trim and Fill procedure did not change the assessment of the relative merits of CBT and PT in reducing weight. Thus, both of these approaches appear to be equally efficacious. For assessing the efficacy of CBT compared to PT (data set #1) and CBT compared to a waiting list (data set #3), both regarding the reduction of *bulimic symptoms* as the outcome of interest, the Trim and Fill procedure actually enlarged the corrected estimates of effect size. Both before and after correction, the efficacy of CBT was assessed as significant greater than zero.

No studies had to be imputed in the three data sets #4, #5 and #6, comparing the efficacy of PT to WL (data set #4) regarding *bulimic symptoms*, and that of CBT compared to PT regarding *psychiatric symptoms* (data set #5), and regarding *interpersonal functioning* (data set #6).

For the data sets #7 to #11, the mean effect size estimates were calculated as *RR*. Data set #7 compared CBT to PT regarding *remission rates,* data set #8 regarding *dropout rates*. Two data sets compared CBT to WL, again regarding *remission rates* (data set #9) and *dropout rates* (data set #10); and one data set compared PT to WL for *dropout rates* (data set #11). The data set of CBT compared to WL regarding remission rates (data set #9) was not affected by publication bias (see Table 1). In contrast, the data set regarding *remission rates* for the comparison of CBT versus PT (data set #7) and all data sets regarding *dropout rates* (data sets #8, #10 and #11) were corrected by the Trim and Fill procedure. In all four of these asymmetric data sets, there was no significant difference between the original and the corrected effect size estimates. As expected, all mean effect estimates were reduced if a

correction was necessary. In the three data sets #8, #10 and #11 both the original and the newly estimated corrected pooled effect sizes did not significantly differ from zero and thus, the evaluation of the efficacy of these interventions was not changed by taking a possible publication bias into account. The efficacy of CBT and PT did not differ in reducing dropout rates (#8), and neither CBT nor PT were more efficacious in reducing dropout rates than WL (#10 & #11). However, the correction according to the Trim and Fill procedure changed the assessment of the efficacy of CBT compared to PT (#7) in reducing remission rates: the effect size turned significant after reduction. The original effect size demonstrated an equal efficacy of both approaches, whereas the corrected effect size indicated a higher efficacy of CBT.

Three data sets estimated the efficacy of prevention programs for universal and high risk samples taken together (28). Two data sets regarding the outcome measures *thin-ideal internalization* (data set #12) and *negative affect* (data set #13), for which the mean effect sizes were calculated as correlation coefficients (*r*), were corrected for publication bias by the Trim and Fill procedure, and missing studies were imputed to compute corrected effect sizes. Both mean effect size estimates were reduced by this procedure, indicating that four primary studies with non-significant effects for prevention programs were missing in each data set. Again, in both data set #12 and #13, there was no significant difference between the original and the corrected estimates of effect size. Confidence intervals for the original effect size estimates were not reported in Stice and Shaw (28), but the effect estimates we replicated, both before and after correction, were significant, as zero was not included in any of the confidence intervals. In conclusion, the correction did not lead to a significant reduction, and the efficacy of prevention programs for changing *thin-ideal internalization* (data set #12) and *negative affect* (data set #13) remained statistically significant. Data set #14, regarding the efficacy of prevention programs in reducing *body mass*, needed no correction according to the Trim and Fill procedure.

To illustrate the Trim and Fill method, Figure 1 shows the funnel plot for the reduction of *overall negative affect at follow up*, the data set with the largest number of studies estimated as missing (N = 4; data set #13).

Insert Figure 1 about here

Results of the rank correlation and regression method

Finally, we assessed the impact of publication bias using the rank correlation (39) and regression methods (40). Both one-sided and two-sided significance tests were conducted. The rank correlation method indicated a significant relationship between effect sizes and their variance in three data sets for one-sided testing. This included *remission rates* in the comparison of CBT and WL (data set # 9), as well as *thin-ideal internalization* (data set #12) and *negative affect* (data set #13) in the data sets reporting the efficacy of prevention programs. In the case of *thin-ideal internalization*, the correlation between effect sizes and their variance was no longer significant if a two-sided test was used. The linear regression method indicated significant asymmetry for the same three effect sizes, using either a one-sided or two-sided significance test. Thus, Begg´s and Egger´s methods led to essentially the same conclusions (see Table 1).

Concordance between the three methods

In 11 of the 14 data sets, none of the methods employed indicated the presence of publication bias. Begg's and Egger's tests were not significant in any of the 11 data sets, and none of the reductions of the Trim and Fill procedure was significant. Thus, there was an almost complete concordance between all three methods in the vast majority of cases.

Begg's and Egger's tests indicated a significant bias in two (three) of the remaining data sets if they were conducted two-tailed (one-tailed). When the Trim and Fill procedure identified the studies missing in these three data sets, their inclusion changed the pattern of

significance only once: the effect size changed from non-significance to significance. Beside this notable exception, in the few data sets in which some evidence for publication bias was found, the evaluation of the efficacy of psychotherapy for eating disorders and of prevention programs was not changed after correcting for a minor tendency toward the selective publication of positive results. Thus, taken together, we found only little evidence for the presence of publication bias. In summary, we can therefore conclude that in most of the data sets where some evidence for publication bias was found, the efficacy of the interventions was unchanged by the selective publication of positive results, and even in the data set where it was changed, the efficacy of CBT remained quite substantial.

DISCUSSION

This study was designed to test for the presence of publication bias in meta-analyses of the efficacy of psychotherapeutic interventions and prevention programs for eating disorders. This was done to identify the most efficacious therapies undistorted by the possible selective publication of positive results for specific treatments.

Depending on the type of test, Begg´s rank correlation method indicated significant publication bias in 3 (21.43 %, one-sided test) or 2 (14.29 %, two-sided test) of the data sets. Regardless of the type of test, Egger´s regression analysis found evidence for publication bias in three of the data sets (21.43 %). The additional significant finding for Egger´s regression analysis might reflect the larger statistical power of this approach (40; 50). Thus, there was some evidence for a selective publication of positive results. Similarly, the Trim and Fill procedure also suggested that there were missing studies in 9 out of 14 (64.29%) data sets reported in the two meta-analyses that qualified for inclusion. However, all of the nine effect sizes were altered non-significantly, as there was never a significant difference between the original and the corrected effect sizes, respectively. In addition, in most of the cases the general pattern of the efficacy of the interventions was not changed by taking a putative

publication bias into account, as can be seen from the fact that both the original and the newly estimated corrected pooled effect sizes were significant in 4 of the 9 data sets in which studies had to be imputed. Both before and after correction, the efficacy of these therapeutic approaches was assessed as significantly larger than zero. Furthermore, for 4 data sets, the original and the corrected effect size estimates were both non-significant. In one data set, however, publication bias changed the pattern of efficacy, as the original RR did marginally fail to reach significance, and the corrected RR turned significant. Therefore, the imputation of presumably missing studies resulted in a significant change of the effect size estimates only once. This result is most important: the impact of publication bias is only once severe and does otherwise not affect the efficacy of the treatments. After correcting for bias, CBT and PT were no longer equally efficacious, as CBT turned out to be superior.

In conclusion, there was thus very little evidence for a selective publication of positive results. The validity of the data sets that were tested and found to be not significantly affected by publication bias is enhanced. We can conclude that the minor tendency toward selective publication of positive results hardly changes the assessment of the efficacy of the therapeutic and preventive interventions.

Interestingly, the two meta-analyses by Hay et al. (23) and Stice and Shaw (28) that provided the data sets for the present analysis both commendably also included unpublished studies and doctoral theses that had not been published in journals. The inclusion of these studies may have contributed to the relatively slight tendency towards publication bias we found, emphasizing the importance of attempting to also include unpublished studies in meta-analyses.

To summarize the efficacy of the treatments under investigation, CBT can be considered efficacious for reducing bulimic symptoms and remission rates, and CBT and PT are equally efficacious in reducing overweight and enhancing interpersonal functioning. CBT

is of higher efficacy than PT in reducing remission rates. However, there is no difference between CBT and PT in reducing dropout rates, and if CBT and PT are compared to WL for this outcome measure of interest, both effect sizes are non-significant, indicating that both treatments are no more efficacious than WL. Furthermore, prevention programs are efficacious in reducing risk factors for the development of eating disorders, namely thin-ideal internalization, negative affect and body mass. To conclude, treatments for eating disorders can mostly be considered efficacious, with the exception of reducing dropout rates.

Limitations of this study

One limitation of the present study was that for some meta-analyses, the necessary raw data needed for a post-hoc assessment of publication bias was no longer available. We therefore recommend that an assessment of small-study and publication bias should become the methodological standard, and sensitivity analyses for publication bias should routinely be included in every meta-analysis, rather than be left to follow-up analyses by other researchers (49; 53; 54). Statistical guidelines such as the "Meta-analysis reporting standards" (MARS) included in the 6th edition of the APA Publication Manual (67; 68) and the "Preferred reporting items for systematic reviews and meta-analyses" (PRISMA; 69) do now include a recommendation to always test for publication bias in meta-analyses.

At present, cognitive behavioral therapy is the only therapy for which a sufficient number of studies exist that allow for a differential evaluation in meta-analyses with regard to different outcome measures. For several other therapies, including interpersonal therapy, hypno-behavioral therapy, dialectic behavioral therapy, supportive therapy and any other psychotherapy such as behavioral weight loss treatment for overweight binge eaters, studies had to be combined in the meta-analysis of Hay et al. (23) in order to obtain sufficiently large data sets. Thus, no differential recommendations regarding their application can be given for these therapies on the grounds of the evidence at hand. Moreover, no meta-analyses could be

retrieved that calculated differential effect sizes for psychoanalysis, humanistic therapy or therapies such as interpersonal or systemic therapy. It thus seems that such studies have either not been conducted, or have not been published. This unfortunate lack of research makes it impossible at present to accurately assess the efficacy of these therapies, and we therefore recommend that additional studies investigating these therapeutic approaches be carried out. Another blind spot of the present body of research is that no pooled effect sizes are available regarding the treatment of AN. There therefore seems to be need for additional studies aimed at identifying treatments for this specific disorder.

From a sociological point of view, a researcher's decision to submit the results of a study for publication is guided by the rules of the social system of the scientific community (70) and by gratifications provided by the scientific system (71; 72). It is often difficult to get negative results published, and researchers may therefore be tempted to invest their effort in potentially more rewarding projects. Similarly, editors may be inclined to give precedence to significant rather than non-significant results. It may well be argued, however, that to avoid producing a biased body of published research, the peer review process should try to identify and reject erroneous results rather than results that are at odds with a researcher's hypothesis, or that are not statistically significant.

A convincing solution to prevent publication bias is the compulsory implementation of a systematic, comprehensive, and freely accessible register of all studies that are conducted in a given field of research. Studies included in such a register can no longer disappear and be left out in future meta-analyses, and the data may even be listed directly in such registers (14; 43). Ultimately, it will only be in domains for which registers have not yet been implemented that a researcher will have to resort to methods such as the Trim and Fill procedure, Begg and Mazumdar´s rank correlation method, and Egger´s regression analysis to assess publication bias (49; 50; 54).

The positive conclusion that can be drawn from the present analysis, however, is that in spite of the current lack of such a central register, publication bias has hardly invalidated the results of prior meta-analyses of the efficacy of psychotherapeutic approaches to treat and prevent eating disorders. The evaluation of the efficacy was only changed for one data set, in favor of CBT. Efficacious prevention programs and treatments are available, and their efficacy is not the result of a reporting bias.

References

Meta-analyses are marked with an asterisk (*)

1. Hoek, H.W. (2006). Incidence, prevalence and mortality of anorexia nervosa and other eating disorders. *Current Opinion in Psychiatry*, *19*, 389-394.

2. Legenbauer, T. & Vocks, S. (2006). *Manual der kognitiven Verhaltenstherapie bei Anorexie und Bulimie* [Manual of cognitive-behavioural therapy for anorexia and bulimia nervosa]. Heidelberg: Springer Medizin Verlag.

3. Brownley, K.A., Berkman, N.D., Sedway, J.A., Lohr, K.N. & Bulik, C.M. (2007). Binge eating disorder treatment: A systematic review of randomized controlled trials. *International Journal of Eating Disorders*, *40*, 337-348.

4. *Vocks, S., Tuschen-Caffier, B., Pietrowsky, R., Rustenbach, S. J., Kersting, A. & Herpertz, S. (2010). Meta-analysis of the effectiveness of psychological and pharmacological treatments for binge eating disorder. *International Journal of Eating Disorders*, *43*. 205-217.

5. Fairburn, C.G. (1997). Eating disorders. In: Clark, D.M. & Fairburn, C.G. (Eds.), *Science and practice of cognitive behaviour therapy* (pp. 209–241). London: Oxford University Press.

6. Hudson, J.L.; Hudson, R.A. & Pope, H. jr. (2005). Psychiatric comorbidity and eating disorders. In: Wonderlich, S.; Mitchell, J.; Zwaan, M. de & Steiger, H. (Eds.), *Eating disorders review, part 1* (pp. 43-58). Radcliff, Oxford.

7. De Zwaan, M. (2001). Binge eating disorder and obesity. *International Journal of Obesity*, *25 (Suppl. 1)*, 51–55.

8. American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders DSM-IV (4th ed.).* Washington, DC: American Psychiatric Association.

9. Lewinsohn, P.M.; Striegel-Moore, R.H. & Seeley, J.R. (2000). Epidemiology and natural course of eating disorders in young women from adolescence to young adulthood. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 1284 – 1292.

10. LeGrange, D. & Lock, J. (2005). The dearth of psychological treatment studies for anorexia nervosa. *International Journal of Eating Disorders*, *37*, 79-91.

11. Jacobi, C., Thiel, A. & Paul, T. (2000). *Kognitive Verhaltenstherapie bei Anorexia und Bulimia Nervosa*. Beltz, Weinheim.

12. Fichter, M.M.; Quadflieg, N. & Gnutzmann, A. (1998). Binge eating disorder: Treatment outcome over a 6-year course. *Journal of Psychosomatic Research*, *44*, 385-405.

13. *Laessle, R. G., Zoettl, C. & Pirke, K.M. (1987). Meta-analysis of treatment studies for bulimia. *International Journal of Eating Disorders*, *6,* 647-653.

14. Rustenbach, S.J. (2003). *Metaanalyse. Eine anwendungsorientierte Einführung* [Meta-analysis. An application-oriented introduction]. Verlag Hans Huber, Bern.

15. Lipsey, M.W. & Wilson, D.B. (2001). *Practical meta-analysis*. Thousand Oaks: Sage Publications Ltd.

16. Sterling, T.D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance- or vice versa. *Journal of the American Statistical Association*, *54,* 30-34.

17. Smith, M.L. (1980). Publication bias and meta-analysis. *Evaluation in Education: An International Review Series, 4,* 22-24.

18. Dickersin, K. (2005). Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 11-34). West Sussex: Wiley & Sons Ltd.

19. Rothstein, H.R., Sutton, A.J. & Borenstein, M. (2005). Publication bias in meta-analysis. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 1-8). West Sussex: Wiley & Sons Ltd.

20. *Cororve Fingeret, M.; Warren, C.S.; Cepeda-Benito, A. & Gleaves, D.H. (2006). Eating disorder prevention research: A meta-analysis. *Eating Disorders*, *14,* 191-213.

21. *Fettes, P. A., & Peters, J.M. (1992). A meta-analysis of group treatments for bulimia nervosa. *International Journal of Eating Disorders*, *11*, 97-110.

22. * Ghaderi, A. & Andersson, G. (1999). Meta-analysis of CBT for bulimia nervosa: Investigating the effects using DSM-III-R and DSM-IVcriteria. *Scandinavian Journal of Behaviour Therapy*, *28*, 79-87.

23. *Hay, P.P.J.; Bacaltchuk, J.; Stefano, S. &Kashyap, P. (2009). Psychological treatments for bulimia nervosa and binging (Review). *The Cochrane Library*, 4.

24. *Hartmann, A., Herzog, T. & Drinkmann, A. (1992). Psychotherapy of bulimia nervosa: What is effective? A meta-analysis. *Journal of Psychosomatic Research*, *36*, 159-167.

25. *Jacobi, C., Dahme, B. & Rustenbach, R. (1997). Vergleich kontrollierter Psycho- und Pharmakotherapiestudien bei Bulimia und Anorexia nervosa. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, *47*, 346-364.

26. * Lewandowski, L. M., Gebing, T. A., Anthony, J. L. & O´Brien, W. H. (1997). Meta-analysis of cognitive-behavioral treatment studies for bulimia. *Clinical Psychology Review*, *17*, 703-718.

27. *Newton, M. S. & Ciliska, D. (2006). Internet-based innovations for the prevention of eating disorders: A systematic review. *Eating Disorders*, *14,* 365-384.

28. *Stice, E. & Shaw, H. (2004). Eating disorder prevention programs: A meta-analytic review. *Psychological Bulletin*, *130*, 206-227.

29. *Stice, E.; Shaw, H. & Marti, C.N. (2007). A meta-analytic review of eating disorder prevention programs: Encouraging findings. *Annual Review of Clinical Psychology*, *3,* 207-231.

30. *Stice, E.; Shaw, H.; Black Becker, C. & Rohde, P. (2008). Dissonance-based interventions for the prevention of eating disorders: Using persuasion principles to promote health. *PrevSci*, *9,* 114-128.

31. * Oesterheld, J.R; McKenna, M.S. & Gould, N.B. (1987). Group psychotherapy of bulimia: A critical review. *International Journal of Group Psychotherapy, 37,* 163-184.

32. * Thompson-Brenner, H., Glass S. & Westen, D. (2003). A multidimensional meta-analysis of psychotherapy for bulimia nervosa.*Clinical Psychology: Science and Practice, 10,* 269-287.

33. * Whitbread, J. & Mcgown, A. (1994). The treatment of bulimia nervosa. What is effective? A meta-analysis. *Indian Journal of Clinical Psychology*, *21*, 32-44.

34. * Whittal, M. L., Agras, W. S. & Gould, R. A. (1999). Bulimia nervosa: A meta-analysis of psychosocial and pharmacological treatments. *Behavior Therapy*, *30*, 117-135.

35. Herzog, T. & Hartmann, A. (1997). Psychoanalytisch orientierte Behandlung der Anorexia Nervosa [Psychoanalytically oriented treatment of anorexia nervosa]. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, *47*, 299-315.

36. Herzog, T. (2000). Stand der vergleichenden Therapieforschung bei Anorexia Nervosa. Ergebnisse einer systematischen Literaturübersicht. *Essstörungen: neue Erkenntnisse und Forschungsperspektiven*, pp. 127-155.

37. Taylor S., Tweedie R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95,* 89-98.

38. Taylor S., Tweedie R (2000b). Trim and Fill: A simple funnel-plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455-63.

39. Begg, C.B. & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*, 1088-101.

40. Egger M, Davey Smith G, Schneider M, Minder C (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629-34.

41. Sterne, J.A.C., Becker, B.J. & Egger, M. (2005). The funnel plot. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 75-98). West Sussex: Wiley & Sons Ltd.

42. Kraemer, H.C., Gardner, C., Brooks III, J.O. & Yesavage, J.A. (1998). Advantages of excluding underpowered studies in meta-analysis: Inclusionist versus exclusionist viewpoints. *Psychological Methods*, *3*, 23-31.

43. Berlin, J.A. & Ghersi, D. (2005). Preventing publication bias: Registries and prospective meta-analysis. In: Rothstein, H.R., Sutton A.J., Borenstein M. (Eds.), *Publication bias in meta-analysis* (pp. 35-48). West Sussex: Wiley.

44. Moshagen, M. & Musch, J. (2008). Publication bias in studies on the efficacy of hypnosis as a therapeutic tool. *Contemporary Hypnosis*, *25*, 94-99.

45. American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders (3rd ed. rev).* Washington, DC: American Psychiatric Association.

46. American Psychiatric Association (2000). *Diagnostical and statistical manual of mental disorders: DSM-IV-TR (4th ed., rev).* Washington: DC: American Psychiatric Association.

47. Degkwitz, R., Helmchen, H., Kockott, G. &Mombour, W., Hrsg. (1980). *Diagnosenschlüssel und Glossar psychiatrischer Krankheiten: dt. Ausgabe d. internat. Klassifikation d. Krankheiten d. WHO, ICD, 9. Revision* [International Classification of Diseases, Ninth Revision]. Chapter V. Springer, Berlin, Heidelberg, New York.

48. World Health Organisation (1992). *The ICD-10 classification of mental and behavioral disorders: Clinical descriptions and diagnostic guidelines*. Geneva: Author.

49. Borenstein M., Hedges L.V., Higgins J.P.T., Rothstein H.R. (2009). *Introduction to meta-analysis*. West Sussex: Wiley & Sons.

50. Sterne, J.A., Gavaghan, D. & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology, 53,* 1119-1129.

51. Terrin N, Schmid CH, Lau J, Olkin I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine*, *22*, 2113–2126.

52. Ioannidis, J.P.A. & Trikalinos, T.A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *Canadian Medical Association Journal*, *176,* 1091-1096.

53. Light, R.J. & Pillemer, D.B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

54. Sterne, J.A.C. & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 99-110). West Sussex: Wiley & Sons Ltd.

55. Cuijpers, P., Smit, F., Bohlmeijer, E., Hollon, S.D. & Andersson, G. (2010). Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. *The British Journal of Psychiatry*, *196*, 173-178.

56. Duval, S. (2005). The Trim and Fill method. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 127-144). West Sussex: Wiley & Sons Ltd.

57. Sutton, A.J., Duval, S.J., Tweedie, R.L., Abrams, K.R. & Jones, D.R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *British Medical Journal*, *320,* 1574-1577.

58. Borenstein M., Hedges L.V., Higgins J.P.T., Rothstein H.R. (2005). *Comprehensive Meta-Analysis, Version 2.2.* Englewood, NJ: Biostat.

59. Rosenthal, R. (1979). The „file drawer problem“ and tolerance for null results. *Psychological Bulletin*, *86*, 638-641.

60. Becker, B.J. (2005). Failsafe N or the file-drawer number. In: Rothstein, H.R., Sutton A.J., Borenstein M. (Eds.), *Publication bias in meta-analysis* (pp. 111-126). West Sussex: Wiley.

61. Hedges, L.V. & Vevea, J. (2005). Selection method approaches. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 145-174). West Sussex: Wiley & Sons Ltd.

62. Begg, C.B. & Berlin, J.A. (1988). Publication bias: A problem in interpreting medical data (with discussion). *Journal of the Royal Statistical Society*, *Series A, 151*, 419-463.

63. Orwin, R.G. (1983). A fail-safe N for the effect size in meta-analysis. *Journal of Educational Statistics*, *8,* 157-159.

64. Fisher, R.A. (1932). *Statistical methods for research workers* (4th Ed.). London: Oliver & Boyd.

65. Gleser, L.J. & Olkin, I. (1996). Models for estimating the number of unpublished studies. *Statistics in Medicine*, *15*, 2493-2507.

66. Hedges, L.V. & Vevea, J. (1996). Estimating effect size under publication bias: Small sample properties and robustness of a random effects selection model. *Journal of Educational and Behavioral Statistics, 21*, 299-332.

67. American Psychological Association (2009). *Publication manual of the American Psychological Association (6th ed.)*. Washington: DC American Psychiatric Association.

68. APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008). Reportings standards for research in psychology. Why do we need them? What might they be? *American Psychologist*, *43,* 839-851.

69. Moher, D., Liberati, A., Tetzlaff, J. & Altman, D.G. (2009). Preferred reporting items for systematic reviews and meta-analysis: The PRISMA statement. *Annals of Internal Medicine*, *151*, 264 - 270.

70. Weingart, P. (2003). *Wissenschaftssoziologie* [Sociology of science]. Bielefeld: transcript Verlag.

71. Luhmann, N. (1970). Die Selbststeuerung der Wissenschaft. In: *Soziologische Aufklärung* [Sociological Enlightenment], *Bd. 5* (pp. 232-252). Opladen: Westdeutscher Verlag.

72. Luhmann, N. (1990). *Die Wissenschaft der Gesellschaft* [The science of society].

Frankfurt/M.: Suhrkamp.

Table 1: Results of the Trim and Fill procedure, Begg and Mazumdar's (1994) rank-correlation test, and Egger's et al. (1997) regression analysis.

Data set number	Author	Intervention/ dependent measure/ Comparison	Original effect size (95% CI)	Replicated effect size (95% CI)	Effect size corrected for publication bias (95% CI)	Number of included studies	Number of missing studies according to Trim and Fill	Begg (Tau) and Egger (β_0)
1	Hay et al. (2009)	CBT vs. PT/ bulimic symptoms/ post-treatment	d = 0.21[a,b] (0.09; 0.34)	0.21[a] (0.06; 0.36)	0.24[a] (0.08; 0.41)	15	1	Tau = .07 β_0 = .7
2	Hay et al. (2009)	CBT vs. PT/ weight/ post-treatment	d = - 0.18[c] (- 0.34; 0.01)	- 0.18 (- 0.34; 0.01)	- 0.12 (- 0.27; 0.03)	11	3	Tau = .35 β_0 = 1.27
3	Hay et al. (2009)	CBT vs. WL/ bulimic symptoms/ post-treatment	d = 0.94[a] (0.7; 1.19)	0.94[a] (0.7; 1.18)	0.98[a] (0.74; 1.22)	12	1	Tau = -.15 β_0 = -1.91
4	Hay et al. (2009)	PT vs. WL/ bulimic symptoms/ post-treatment	d = 1.14[a,b] (0.89; 1.39)	1.14[a] (0.89; 1.38)		7	0	Tau = -.47 β_0 = -2.11
5	Hay et al. (2009)	CBT vs. PT/ general psychiatric symptoms/ post-treatment	d = 0.13[b] (-0.09; 0.35)	0.14 (-0.08; 0.37)		7	0	Tau = -.33 β_0 = -1.71

Data set number	Author	Intervention/ dependent measure/ Comparison	Original effect size (95% CI)	Replicated effect size (95% CI)	Effect size corrected for publication bias (95% CI)	Number of included studies	Number of missing studiesaccor-ding to Trim and Fill	Begg (Tau) and Egger (β_0)
6	Hay et al. (2009)	CBT vs. PT/ interpersonal functioning/ post-treatment	d = 0.12[b] (-0.05; 0.28)	0.12 (-0.04; 0.28)		7	0	Tau = -.05 β_0 = .37
7	Hay et al. (2009)	CBT vs. PT/ remission/ post-treatment	RR = 0.87 (0.74; 1.02)	0.87 (0.74; 1.02)	0.79[a] (0.65; 0.95)	10	2	Tau = .02 β_0 = .74
8	Hay et al. (2009)	CBT vs. PT/ drop out due to any reason/ post-treatment	RR = 0.97 (0.70; 1.35)	0.97 (0.70; 1.35)	0.84 (0.60; 1.16)	14	3	Tau = -.13 β_0= -.60
9	Hay et al. (2009)	CBT vs. WL/ remission/ post-treatment	RR = 0.69[a] (0.61; 0.79)	0.69[a] (0.61; 0.79)		8	0	Tau = -.64*† β_0 = -3.53 *†
10	Hay et al. (2009)	CBT vs. WL/ drop out due to any reason/ post-treatment	RR = 1.46 (0.77; 2.79)	1.45 (0.78; 2.71)	1.27 (0.66; 2.44)	11	1	Tau = .29 β_0= 1.26
11	Hay et al. (2009)	PT vs. WL/ drop out due to any reason/post-treatment	RR = 1.44 (0.83; 2.49)	1.44 (0.83; 2.49)	0.94 (0.49; 1.8)	6	3	Tau = .07 β0 = .87

Data set number	Author	Intervention/ dependent measure/ Comparison	Original effect size (95% CI)	Replicated effect size (95% CI)	Effect size corrected for publication bias (95% CI)	Number of included studies	Number of missing studiesaccor-ding to Trim and Fill	Begg (Tau) and Egger (β_0)
12	Stice & Shaw (2004)	Thin-ideal internalization /overall/ follow up	r = 0.15 (-)	0.12[a] (0.07; 0.16)	0.10[a] (0.05; 0.14)	23	4	Tau = .35* β_0 = 1.42*†
13	Stice & Shaw (2004)	Negative affect/overall/ follow up	r = 0.09 (-)	0.08[a] (0.04; 0.12)	0.06[a] (0.02; 0.11)	17	4	Tau = .43*† β_0 = 1.28*†
14	Stice & Shaw (2004)	Body mass/ overall/ follow up	r = 0.05 (-)	0.09[a] (0.04; 0.13)		11	0	Tau = -.04 β_0 = -.31

Note. All effect sizes are based on the random effects model.CBT = Cognitive behavioral therapy; CI = confidence interval; d = standardized mean difference; PT = psychotherapy; RR = relative risk; r = correlation; WL = waitlist.

[a] Significant effect size with a confidence interval not including zero.

[b] The original minus sign for this effect size was converted into a positive sign to make effect sizes larger than 0 indicate the efficacy of a treatment.

[c] The original positive sign for this effect size was converted into a negative sign to make effect sizes larger than 0 indicate the efficacy of a treatment.

* p< .05, one-sided; †p< .05, two-sided

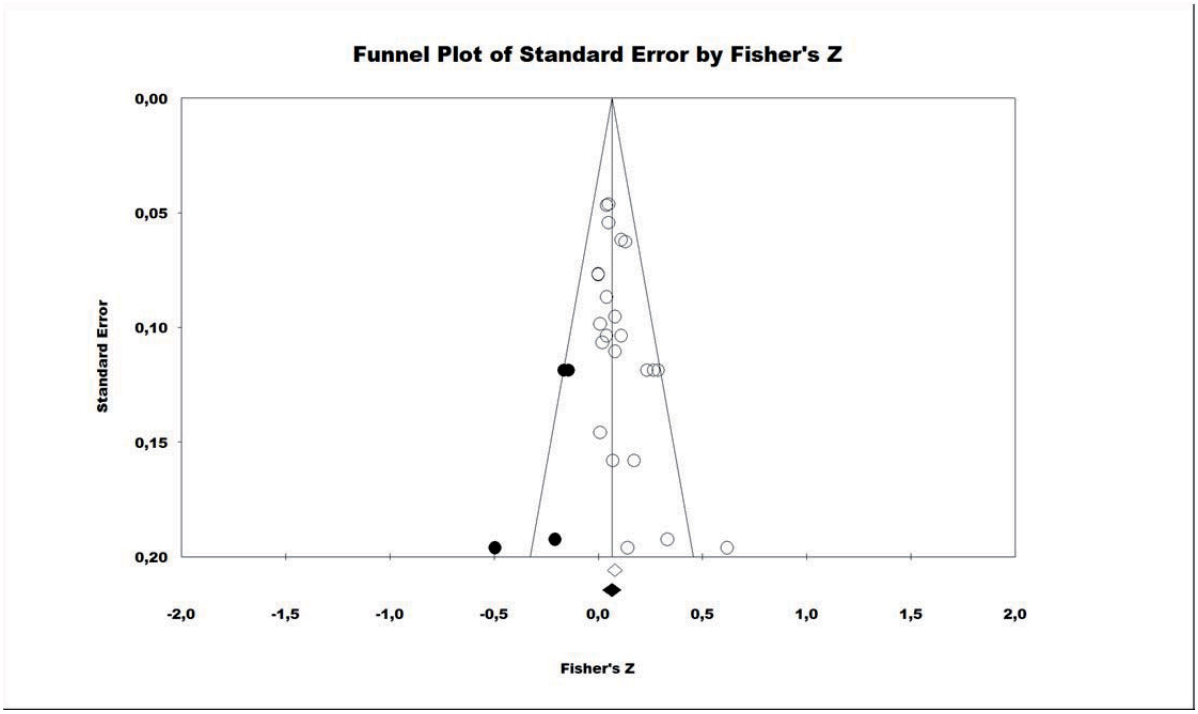

Figure 1.

Note. An illustration of the Trim and Fill method showing the funnel plot for the reduction of overall negative affect after delivery of prevention programs at follow up, displaying Fisher's z (data set #13; Stice & Shaw, 2004). The dark circles indicate the four studies identified as missing.

18.10.11

Dear Mrs. Helen Niemeyer,

Your submission entitled "Publication Bias in Meta-Analyses of the Efficacy of Psychotherapeutic Interventions for Schizophrenia" has been received by Schizophrenia Research

You may check on the progress of your paper by logging on to the Elsevier Editorial System as an author. The URL is http://ees.elsevier.com/schres/.

Your username is: HNiemeyer

If you need to retrieve password details, please go to:
http://ees.elsevier.com/schres/automail_query.asp

Your manuscript will be given a reference number once an Editor has been assigned.

Thank you for submitting your work to this journal.

Kind regards,

Elsevier Editorial System

Publication Bias in Meta-Analyses of the Efficacy of Psychotherapeutic Interventions for Schizophrenia

Helen Niemeyer, Jochen Musch, Reinhard Pietrowsky

Heinrich-Heine-University, Department of Experimental Psychology, Universitätsstraße 1, 40225 Düsseldorf, Germany.

Corresponding author:

Helen Niemeyer, Heinrich-Heine-University, Department of Experimental Psychology, Universitätsstraße 1, 40225 Düsseldorf, Germany; Tel: ++49 211 81-12272; FAX: ++49 211 81-14261; Email: helen.niemeyer@hhu.de

None of the authors has any conflict of interest.

Abstract

Objective: Meta-analyses are prone to publication bias, the problem of selective publication of studies with positive results. It is unclear whether the efficacy of psychotherapeutic interventions for schizophrenia is overestimated due to this problem. This study aims at enhancing the validity of the results of meta-analyses by investigating the degree and impact of publication bias.

Method: Begg and Mazumdar´s adjusted rank correlation test, Egger´s regression analysis and the Trim and Fill procedure were applied to all systematic reviews up to September 2010 that reported the necessary data to assess publication bias.

Results: We examined 22 data sets, reported in 10 meta-analyses, for indications of publication bias. Begg's test indicated significant bias in 2 (9.09%) of these data sets, while Egger's test found bias in 3 (13.64%) of the data sets. The correction by the Trim and Fill procedure changed the significance of an effect size only once (4.55%), and did so unexpectedly in favor of the treatment condition. Psychosocial family interventions, regarding the outcome measure "leaving study early" in the time period between 13 and 24 months, were shown to be efficacious.

Discussion: Overall, we found only moderate evidence for the presence of publication bias. With one notable exception, the pattern of efficacy of psychotherapy for schizophrenia was not changed in the data sets in which publication bias was found. Several efficacious therapies exist, and their efficacy does not seem to be the result of publication bias.

Keywords

Schizophrenia, publication bias, meta-analysis, psychotherapy research

1. Introduction

The efficacy of psychotherapeutic interventions to treat schizophrenia, a serious mental illness with a median lifetime prevalence of 4 per 1000 persons (McGrath, Saha, Chant & Welham, 2008), has been investigated in a large number of empirical studies. In order to obtain an overview over this multitude of studies and to retrieve results with increased inferential power and validity, meta-analyses have been used to synthesize these results. Being based upon bigger samples, meta-analyses provide comprehensive and more precise evidence for the efficacy of therapeutic interventions (Rustenbach, 2003). Psychotherapeutic and cognitive interventions for schizophrenia which have frequently been investigated in meta-analyses comprise cognitive behavioral therapy (CBT), family oriented therapy, integrated psychological therapy (IPT), psychoeducation, and cognitive remediation trainings. The effect size estimates for these interventions provide an evidence base for guidelines and recommendations for clinical practice. On the basis of such empirical results, the official treatment guidelines in both the United Kingdom and the United States regard psychological interventions as essential for the treatment of schizophrenia, and now include CBT, especially for psychosis (NICE, 2003, 2009).

However, the validity of the results of meta-analyses dealing with the efficacy of psychotherapeutic interventions may be affected by selective reporting of positive outcomes-publication bias (Rothstein, Sutton & Borenstein, 2005). Publication bias is a systematic bias that is characterized by the preferential publication of studies with significantly positive results, as opposed to null or non-significant results (Hopewell, Clarke & Mallett, 2005). If the magnitude and direction of an effect influence the decision to submit or accept a research report, a positivity bias is induced in the published body of research. As unpublished studies

are not randomly missing, but are systematically omitted due to the non-significance or negativity of their results, reviews will not accurately reflect of the totality of knowledge (Rustenbach, 2003). The validity of the results of a meta-analysis is highly dependent on the comprehensiveness of the underlying data, and if only published studies are included, the efficacy of psychotherapeutic interventions may be overestimated, whereas the inclusion of unpublished studies with non-significant results may reduce the mean effect estimate (Hopewell et al., 2005; Lipsey & Wilson, 2001; Rothstein et al., 2005; Rustenbach, 2003).

To investigate how thoroughly publication bias was assessed in meta-analyses about the efficacy of psychotherapeutic interventions for schizophrenia, and how exhaustively unpublished studies were included, we conducted a comprehensive literature search to identify all meta-analyses that have been published up to September 2010. We located 46 meta-analyses in which the efficacy of CBT, psychoeducation, integrated psychological therapy (IPT), cognitive remediation training, family-oriented interventions, social skills training, token systems, psychodynamic therapy, psychoanalysis, hypnosis or music therapy was investigated (all meta-analyses are marked in the list of references with an asterisk). Only 4 of the 46 meta-analyses (8.7%) included unpublished studies (Benton & Schroeder, 1990; Jones, Cormac, Silveira da Mota Neto & Campbell, 2010; Pharoah, Mari, Rathbone & Wong, 2010; Silverman, 2003). The efficacy of the interventions may therefore have been overestimated. Nevertheless, we found that in none of these 46 meta-analyses the Trim and Fill procedure (Taylor & Tweedie, 2000a, 2000b) or Begg and Mazumdar´s rank correlation test (Begg & Mazumdar, 1994) were applied to statistically control for publication bias. Egger´s regression analysis (Egger, Davey Smith, Schneider & Minder, 1997) was applied only once (Villeneuve, Potvin, Lesage & Nicole 2010), and indeed publication bias was found. This emphasizes the need to control for bias in meta-analyses for schizophrenia, but

there appears to be an almost complete absence of statistical control for publication bias. In 6 meta-analyses (Benton & Schroeder, 1990; Kurtz & Mueser, 2008; Lincoln, Wilhelm & Nestoriuc, 2007; Lincoln, Suttner & Nestoriuc, 2008; Mueller, Roder & Brenner, 2007; Roder, Mueller, Mueser & Brenner, 2006; Zimmermann, Favrod, Trieu & Pomini, 2005) a funnel plot was applied, and in 7 analyses (Jones et al., 2010; Mojtabai, Nicholson & Carpenter, 1998; Nose, Barbui, Gray & Tansella, 2003; Pharoah et al., 2010; Villeneuve et al., 2010; Wykes, Steel, Everitt & Tarrier, 2008) failsafe N´s were computed, but due to major drawbacks of these methods, described in detail in the methods section and discussion, these assessments cannot be considered appropriate. In conclusion, meta-analyses of psychotherapeutic interventions for schizophrenia have not been comprehensively assessed for publication bias, and the extent of publication bias in these meta-analyses is unknown.

As clinicians depend on meta-analytic results for deciding on their patients' treatment, an overestimation of the efficacy of therapeutic interventions can have severe effects (Dickersin, 2005). Thus, publication bias is not merely a statistical problem, but can also have dire practical implications. Biased empirical evidence will prompt psychotherapists to apply interventions in routine care that are erroneously seen as being efficacious. As a result, inefficacious psychotherapeutic interventions could even have detrimental effects for patients with schizophrenia.

As we have shown above, no comprehensive statistical assessment of publication bias has hitherto been carried out in the field of psychotherapy research for schizophrenia. Although psychotherapeutic interventions for schizophrenia are claimed to be efficacious, the efficacy may be overestimated due to publication bias, and it might have to be reduced if publication bias would be taken into account. Thus, we undertook an investigation of all

meta-analyses of the efficacy of therapeutic interventions for schizophrenia. The present study aimed at re-analyzing all meta-analyses which investigate the efficacy of psychotherapeutic interventions for schizophrenia, conducted up to September 2010, by means of a statistical assessment of publication bias. We tested for the existence of a possible publication bias and calculated corrected effect size estimates if bias was present, assessing its impact by testing the difference between the original and the unbiased effect sizes for significance.

2. Methods

2.1 Data sources

We applied the search strategies recommended by Lipsey and Wilson (2001) to locate all published meta-analyses about psychotherapeutic interventions for schizophrenia. We conducted a comprehensive search in electronic databases of the literature (PsycInfo and Psyndex), and used cross-references in other articles and book chapters. We applied the keywords “meta-analysis” or “systematic review” and ”schizophrenia”, “schizoaffective disorder” or “delusional disorder” in combination. The inclusion was restricted to meta-analyses reported in English or German, up to September 2010.

2.2 Study selection and data extraction

Articles were retrieved for further assessment if the title or abstract suggested that they dealt with a meta-analysis about psychotherapy for schizophrenia. If an abstract provided insufficient information, the respective article was examined in order not to miss a relevant review. A meta-analysis had to involve studies with schizophrenic samples that had been diagnosed with schizophrenia or schizoaffective disorder according to the diagnostic criteria of either DSM-III-R (APA, 1987), DSM-IV (APA, 1994), DSM-IV-TR (APA, 2000), ICD-9

(Degkwitz, Helmchen, Kockott, & Mombour, 1980) or ICD-10 (World Health Organization, 1992). Inclusion was restricted to reviews that investigated the efficacy of at least one psychotherapeutic intervention, and there were no restrictions on their nature. All meta-analyses examining the efficacy of pharmacological treatment or electroconvulsive treatment were excluded, as these are not psychotherapies. If samples composed of other disorders along with schizophrenia were included in a meta-analysis, and the effect sizes were combined to overall effect estimates not restricted to the treatment of schizophrenia, the respective meta-analysis was excluded.

When several arms of studies were reported in a meta-analysis, we included all arms for which a homogeneous pooled effect size estimate, the primary studies´ effect sizes and a measure of their precision were provided. These data are necessary for the application of the statistical methods for the assessment of publication bias. If the data required were not reported in the paper, we made an attempt to obtain them from the authors. Confidence intervals, standard deviations, standard errors and variances were acceptable measures of precision. If the effect size statistic was a correlation coefficient, the sample size could be used instead. If the primary studies´ effect sizes and a measure of their precision were not provided, but the means and standard deviations or variances of the raw data were given, we used them to re-calculate the effect sizes and their precision and included the respective arm of each meta-analysis for which the effect size could be confirmed. Homogeneity of the pooled effect sizes was also taken into account, since the statistical methods to detect the presence of publication bias are based on a homogeneity assumption (Ioannidis & Trikalinos, 2007). In case of between-study heterogeneity these statistical methods are inappropriate and lead to false alarms (Sterne, Gavaghan & Egger, 2000; Taylor & Tweedie, 2000a, 2000b; Terrin, Schmid, Lau & Olkin, 2003). We thus excluded heterogeneous data sets from the

present study in order to avoid erroneous conclusions. We also excluded all arms of a meta-analysis including five or fewer trials. The application of statistical methods to detect publication bias is unreliable and underpowered if the number of studies in a data set is too small (Sterne, Becker & Egger, 2005). We adopted our cut-off criterion from Egger and colleague's paper, in which the funnel plot method and regression analysis are introduced in detail (Egger et al., 1997).

2.3 Assessment of publication bias

The most familiar methods for the examination of publication bias are based on the funnel plot (Light & Pillemer, 1984). The funnel plot is a graph of the effect sizes of the primary studies included in a meta-analysis on the abscissa, with a measure of their precision plotted on the ordinate (Light, Singer & Willett, 1994). Large studies tend to cluster near the mean effect size at the top of the graph, and smaller studies are more widely dispersed at the bottom of the graph. The plot is shaped like a funnel in the absence of bias, and asymmetry in the dispersion of the studies is seen as an indicator of a publication bias. If there is a selective reporting of significant results, smaller studies are more likely to meet the criterion for statistical significance and hence to be published if they have large effect sizes. Therefore, in the presence of bias a higher concentration of studies will appear on the bottom of the plot on one side of the mean effect estimate (Sterne et al., 2005). The major disadvantage of this method is that the interpretation is largely subjective, because the plot can only be assessed visually. For this reason, formal quantitative tests have been developed to assess publication bias objectively (Sterne & Egger, 2005).

Begg and Mazumdar's rank correlation method (1994) and Egger´s regression analysis (Egger et al., 1997) both investigate whether the effect sizes of the primary studies are

associated with a measure of their precision. Begg and Mazumdar´s rank correlation test utilizes the rank correlation, Kendall´s Tau, between the standardized effect sizes and their variances. In the absence of publication bias, there is no relationship between the effect sizes and their variance, but if publication bias is present a large variance should be associated with a large effect size, since small studies, which have higher variances, are more likely to be published if they show large effect sizes (Sterne & Egger, 2005). Testing the correlation for significance therefore provides a statistical test for small-study effects and hence, publication bias (Begg & Mazumdar, 1994). In Egger´s linear regression method (Egger et al., 1997), the standard normal deviate (the effect size standardized by its standard error) is regressed on its precision (defined as the inverse of its standard error). If the intercept does not pass through the origin, this indicates asymmetry in the dispersion and can be used as measure of the degree of asymmetry. Test power is generally higher than that of the rank correlation test. Sterne and Egger (2005) recommend using two-sided p-values, whereas Cuijpers, Smit, Bohlmeijer, Hollon and Andersson (2010) have argued that the tests should be conducted one-tailed. We therefore decided to report both one- and two-sided significance tests, and set the probability of a type I error at .05.

A third method, Trim and Fill (Taylor & Tweedie, 2000a, 2000b), adds an estimation of the number of missing studies and their outcome to the assessment of the funnel plot´s asymmetry. The number of putatively missing studies is estimated on the basis of the asymmetry of the funnel plot. If there is no bias, the effect sizes are distributed symmetrically around the true effect size. In the case of asymmetry, the number of probably missing studies is estimated by first using an iterative algorithm to trim off the asymmetric right-hand side of the funnel plot, which consists of the largest effect sizes, and then re-filling the trimmed studies and the imputed missing studies with null and negative effects on the left-hand side

around the symmetric remainder. This procedure results in a symmetrical dispersion. After a few iterations, the procedure usually stops because symmetry is reached, and a new combined effect estimate, which is adjusted for bias, as well as the number of probably missing studies are provided. Importantly, the number of studies should not be regarded as the conclusive number of missing studies, but should rather be seen as a sensitivity analysis of how robust the initial effect estimate is in the face of publication bias (Duval, 2005). A corrected 95% confidence interval is computed based on both the observed and the imputed studies. In case of publication bias, we tested the difference between each effect size estimate and its corrected value for significance by examining whether the original estimate remained within the confidence limits of the corrected effect size.

All analyses were conducted using the software "Comprehensive Meta-Analysis" (Version 2.2; Borenstein, Hedges, Higgins & Rothstein, 2005), which was developed for the computation of meta-analytic effects. In the present study, odds ratios (*OR*) and risk ratios (*RR*) smaller than 1 indicate that a treatment is more efficacious in the experimental group.

3. Results

3.1 Description of meta-analyses investigated

Ten meta-analyses on the efficacy of psychotherapeutic interventions for schizophrenia fulfilled all inclusion criteria and qualified for the present re-analysis. Publication bias has not yet been assessed comprehensively in these meta-analyses. All arms of these meta-analyses for which the necessary data was provided and that had resulted in homogeneous effect sizes were included, 22 data sets in all (see Table 1). These 22 data sets fulfilled all of the following inclusion criteria: 1) the efficacy of psychotherapeutic interventions for schizophrenia was assessed; 2) effect size statistics and a measure of

precision for all included primary studies, or the raw data for their calculation, were provided and the pooled effect size estimate could be confirmed; 3) the effect sizes were homogeneous; and 4) at least six primary studies were integrated (Sterne et al., 2005).

Five of the included meta-analyses dealt with the efficacy of CBT (Jones et al., 2010; Lincoln et al., 2008; Lynch, Laws & McKenna, 2010; Wykes et al., 2008; Zimmermann et al., 2005), one with family psychosocial interventions in community settings (Pharoah et al., 2010), one with combined family therapy and psychoeducation, as well as psychoeducation alone (Pitschel-Walz, 1997), and two with family therapy and psychoeducation (Pitschel-Walz & Engel, 1997; Pitschel-Walz, Leucht, Bäuml, Kissling & Engel, 2001). The last meta-analysis (Bola, 2006) investigated the efficacy of treatment without concurrent medication as compared to treatment including medication. A complete list of all 22 data sets is given in Table 1, which also reports the number of studies included in each data set.

3.1.1 Meta-analyses of the efficacy of CBT

In Lynch et al. (2010) the efficacy of CBT was compared to control conditions. Only randomized controlled trials (RCTs) were included; all had been published. Two data sets (#1, #2) fulfilled our inclusion criteria. In Lincoln et al. (2008), the focus was placed on investigating the cognitive core elements of CBT, especially cognitive restructuring. Only published RCTs were included. One data set (data set #3) was included in the present analysis. Wykes et al. (2008) analyzed the effectiveness of CBT compared to control conditions. Only published RCTs were included. However, as the quality of the included studies varied considerably, Wykes et al. (2008) calculated separate effect sizes for studies of high and low quality, and we were able to include one data set in which the studies of low quality were combined (data set #4). Zimmermann et al. (2005) also investigated the efficacy

of CBT compared to control conditions. Only published studies were included, and all had to be RCTs. We included five data sets from this meta-analysis (#5- #9). Jones et al. (2010) compared the efficacy of CBT to standard care. Only RCTs were included. Unpublished studies were explicitly included, and one of those was integrated into the data set (#10) that fulfilled the inclusion criteria for the present study.

3.1.2 Meta-analyses of the efficacy of other psychotherapeutic interventions

Family-oriented interventions, including family therapy with the patient and the family, as well as relatives´ groups or family sessions without the patient, were investigated in the meta-analysis of Pharoah et al. (2010). The family-oriented interventions were compared to standard care, including pharmacological interventions. Randomized and quasi-randomized trials were included. One unpublished study was included, but it was not integrated into any of the three data sets we assessed for publication bias (#11, #12 & #13). Family-oriented interventions including psychoeducation, compared to standard care, were investigated by Pitschel-Walz and colleagues in three meta-analyses (Pitschel-Walz, 1997; Pitschel-Walz & Engel, 1997; Pitschel-Walz et al., 2001). Only published studies were included. In seven data sets (#14 - #19, #21) the efficacy of family therapy including psychoeducation was investigated, and in one data set (20) the efficacy of psychoeducation alone. Subsamples were used in these data sets, such as only studies on short-term interventions, for example. Finally, Bola (2006) investigated the efficacy of psychotherapeutic treatment with initial medication free periods in early episode schizophrenia, as compared to medical treatment or psychosocial treatment plus medication. Only published studies were included. As too few RCTs existed, the inclusion criterion was set to the integration of quasi-experimental studies. We were able to include one data set, #22.

In none of the data sets we included from the ten meta-analyses, comparisons between two efficacious psychotherapeutic treatments were conducted. With the exception of Bola´s analysis of medicated versus unmedicated treatment (2006), all comparisons included control groups such as standard care or TAU.

All of the data sets provided by Lincoln et al. (2008), Wykes et al. (2008) and Zimmermann et al. (2005) refer to post-treatment effect sizes. Jones et al. (2010), Pharoah et al. (2010) and Bola (2006) reported follow-up effect sizes. Pharoah et al. (2010) differentiated between different follow-up periods, namely 7-12 months and 13-24 months. Lynch et al. (2010) and Pitschel-Walz and colleagues (Pitschel-Walz, 1997; Pitschel-Walz & Engel, 1997; Pitschel-Walz et al., 2001) provided both post and follow-up effects sizes in the data sets that qualified for inclusion in the present study. In order to avoid false positives, we decided not to include the effects that were additionally reported in all of the ten meta-analyses because we found these effect sizes to be heterogeneous using a Q-test, offering an obvious alternative explanation for an apparent asymmetry of the respective effect sizes.

Overall, our recalculations performed on the raw data matched the original ones for most of the mean effect sizes. However, some negligible deviations from the original estimates occurred, at most .04 for the mean differences (*d*), .03 for the relative risk (*RR*), and .03 for the correlation coefficients (*r*). Presumably, these slight discrepancies were the result of different rounding. We used the recalculated effect sizes for the statistical assessment of publication bias. All original and recalculated effect sizes are provided in Table 1.

3.2 Results of the Trim and Fill calculations

On the basis of Trim and Fill calculations, from 1 to 5 missing studies were identified in 15 of the 22 data sets: in data set #2 until #8, #12 until #17, as well as #19 and #21. We

calculated corrected mean effect size estimates for all the asymmetric data sets, which were somewhat reduced. Nevertheless, these corrections for publication bias never significantly reduced the efficacy estimates. No studies had to be imputed in the remaining data sets. Table 1 specifies the detailed results of the Trim and Fill calculations for all outcome measures.

Insert Table 1 about here

3.2.1 Results for meta-analyses investigating the efficacy of CBT

No missing studies were found in data set #1 (Lynch et al., 2010), measuring the efficacy of CBT in reducing general symptoms. The original effect size was non-significant, indicating that there is no difference between CBT and the different control groups included. In data set #2 (Lynch et al., 2010), measuring the efficacy of CBT in comparison to control groups in reducing relapse rates, one missing study was imputed. As expected, the newly estimated effect size was reduced. However, the difference between the original and the corrected effect size estimate was not significant. The mean effect size estimate of data set #2 was non-significant before and after the reduction, also indicating that CBT is no more efficacious than the control conditions.

A further data set (data set #3; Lincoln et al., 2008) assessed the efficacy of CBT in reducing general symptoms compared to treatment as usual (TAU). However, despite the imputation of one additional study and reduction of the corrected effect size estimate with the Trim and Fill procedure there was no significant difference between the original and the corrected effect size estimate. Both before and after correction, the effect size estimates differed significantly from zero. Thus, the assessment of the relative merits of CBT and TAU in reducing the general symptoms of schizophrenia was not changed by the reductions carried out with the Trim and Fill procedure.

Wykes et al. (2008) provided a data set regarding the efficacy of CBT in reducing positive symptoms (#4). Two missing studies were imputed by the Trim and Fill procedure and the effect size was reduced, as expected. Again, there was no significant difference between the original and the corrected effect size estimate, and in addition both effect size estimates differed significantly from zero. Thus, selective reporting did not change the overall assessment of the efficacy, which in the case of CBT appears to be substantial.

The meta-analysis of Zimmermann et al. (2005) provided five data sets (#5- #9). Four of these data sets were corrected for publication bias, and missing studies were imputed. As expected, all four pooled effect sizes were reduced. Up to five missing studies were imputed, respectively. Three studies were missing in the data set comparing the efficacy of CBT to TAU in reducing positive symptoms (data set #5). This procedure led to a non-significant reduction of the mean effect size estimate. Five studies with non-significant effects for CBT were imputed in data set #6 also regarding the reduction of positive symptoms, but comparing CBT to all included control conditions. Again, the mean effect size estimate was reduced only non-significantly. The Trim and Fill procedure imputed 4 missing studies in data set #7 regarding the efficacy of CBT compared to all included control conditions in reducing positive symptoms in the subsample of chronic patients, and 2 missing studies in data set #8, investigating the reduction of positive symptoms in the same subsample, but for CBT compared only to TAU. All corrected effect size estimates were significant both before and after the reduction, and thus the efficacy of CBT in reducing positive symptoms remains despite a small tendency for selective reporting of positive outcomes. In data set #9, reporting a significant effect size estimate in favor of CBT, no study was found to be missing.

Jones et al. (2010) found CBT to be more efficacious than standard care (#10), and no study was found to be missing.

3.2.2 Results for meta-analyses of the efficacy of other psychotherapeutic interventions

In order to assess the efficacy of family-oriented interventions compared to standard care, Pharoah et al. (2010) provided three data sets (#11, # 12 & #13). No missing studies were found for data set #11. In contrast, the data set regarding leaving study early (#12), regarding the time period between 7 and 12 months, as well as data set #13, regarding leaving study early in the period between 13 and 24 months, were both corrected by the Trim and Fill procedure, and one study was found to be missing in each of these two data sets. Both effect sizes were reduced, as expected. The effect size estimates in data set #12 were non-significant before and after the reduction, indicating that family interventions are no more efficacious than standard care. In data set #13, however, the mean effect size estimate was almost significant before the reduction, as the upper confidence limit was exactly 1 for the original *RR*, and it changed slightly to an upper limit .98 after the reduction, indicating efficacy for family interventions after the correction for publication bias.

Six data sets estimated the efficacy of a combination of family interventions and psychoeducation as compared to standard care (Pitschel-Walz, 1997; Pitschel-Walz & Engel, 1997; Pitschel-Walz et al., 2001), all regarding relapse rates. Five of these data sets where found to have missing studies. The Trim and Fill procedure estimated 2 missing studies in both data set #14, including samples with merely schizophrenic patients, and data set #17; regarding the efficacy of family interventions and psychoeducation after one year. One missing study each was found in data set #15, including all effect size estimates at all

measurement times, and data set #16, including only studies with short-term interventions. Three missing studies were imputed in data set #21, regarding the efficacy of the treatment in reducing relapse for post and follow-up measurement combined. The reductions were non-significant in all data sets, and all effect size estimates were significant both before and after the reduction. Thus family therapy combined with psychoeducation appears to be efficacious. Data set #18 including 17 studies measuring effects on relapse rates was not affected by publication bias. The robust effect size estimate was found to be significant, indicating that family therapy and psychoeducation combined are efficacious in reducing relapse rates.

According to the Trim and Fill procedure, 2 studies were missing in data set #19, estimating the efficacy of family interventions and psychoeducation compared to standard care, using general symptoms as outcome of interest. As could be expected, the effect size estimate was reduced. Both the original and the reduced effect size estimate differed significantly from zero, indicating the efficacy of the treatment. In summary, all effect sizes reported by Pitschel-Walz and colleagues (Pitschel-Walz, 1997; Pitschel-Walz & Engel, 1997; Pitschel-Walz et al., 2001) assessed in the present study that were affected by publication bias were reduced. The Trim and Fill procedure revealed no publication bias in data set #20, showing higher efficacy of psychoeducation as compared to standard care in reducing relapse rates.

Data set #22 (Bola, 2006), regarding the efficacy of unmedicated compared to medicated treatment for different outcome measures taken together, was not found to be in need of correction by the Trim and Fill procedure. This effect size was non-significant, indicating that both interventions, treatment with and without medication, appear to be equally efficacious.

As a demonstration of the imputation of missing studies by the Trim and Fill procedure, the funnel plot for the data set with the largest number of studies estimated as missing (N = 5; data set #6; Zimmermann et al., 2005) is displayed in Figure 1. The mean effect size estimate for the reduction of positive symptoms through CBT at post-treatment was non-significantly reduced (see Figure 1).

Insert Figure 1 about here

3.3 Results of the rank correlation and regression method

The presence of publication bias was additionally assessed using the rank correlation (Begg & Mazumdar, 1994) and the regression method (Egger et al., 1997). We conducted both one-sided and two-sided significance tests. The rank correlation method yielded significant results for two data sets both under the one-sided and two-sided significance test, indicating a relationship between the effect sizes and their variance. The comparison of CBT and supportive therapy regarding general symptoms of schizophrenia at post-treatment was affected by publication bias (data set #1), as was the comparison of family-based interventions to standard care regarding hospital admission between the 7th and 12th month of treatment (data set #11). The linear regression analysis revealed a significant asymmetry in 4 data sets for one-sided testing, including: data sets #1 and #11 described above, and data sets #5 and #6, regarding the efficacy of CBT compared to TAU (#5) and compared to all control conditions (#6) in reducing positive symptoms at post-treatment. For the 3 data sets #1, #6 and #11 Egger´s method indicated publication bias in two-sided testing, as well. In summary, Begg´s and Egger´s methods led to the same conclusions for 2 data sets (#1 & #11), and Egger´s method indicated publication bias for 2 further data sets (#5 & #6), probably due to its higher power (see Table 1).

3.4 Concordance between the three methods

In the majority of the 22 assessed data sets (18 data sets, 81.82%; data sets #2- 4, 7-10, 12- 22), none of the methods employed indicated the presence of publication bias. Begg's and Egger's tests found no significant correlations in these 18 data sets, and the reduction of the effect size estimates due to the imputations of missing studies, according to the Trim and Fill procedure, never reached significance. Thus, there was an almost complete agreement between the three methods. Only regarding up to 4 data sets, Begg's and Egger's tests indicated a significant bias in 2/3 (2/4) of the data sets if they were conducted two-tailed (one-tailed). When the Trim and Fill procedure identified the studies missing in these data sets, their inclusion changed the pattern of significance only once, and did so even in favor of the treatment condition. Taken together, we found only very little evidence for the presence of publication bias.

4. Discussion

We investigated the possibility that the treatment effects in meta-analyses of psychotherapeutic interventions for schizophrenia are overestimated due to selective publication of studies with significant results. We endeavored to include meta-analyses for all psychotherapeutic interventions, in order to comprehensively assess the field of psychotherapy research for schizophrenia. We assessed the presence of publication bias and its impact on all homogeneous data sets, for which the necessary effect statistics were provided. Our study aimed at identifying the most efficacious therapies for schizophrenic patients, undistorted by a putative publication bias.

Publication bias does exist in psychotherapy research about the efficacy of psychotherapeutic interventions on schizophrenia, but it was found only in a minority of data

sets, and in spite of bias, the pattern of efficacy for most of the reviewed interventions did not change. Of the 22 data sets reported in the 10 meta-analyses that qualified for inclusion, Begg´s rank correlation method indicated significant publication bias in 2 cases (9.09 %; one- and two-tailed). In 4 data sets (18.18 %; one-tailed) or 3 (13.64%; two-tailed) Egger´s regression analysis found evidence for publication bias. The greater bias found by Egger´s regression analysis may reflect the larger statistical power of this approach (Sterne et al., 2000; Sterne & Egger, 2005). These particular data sets need to be treated with caution due to an association between the primary studies´ effect sizes and their measure of precision, namely the comparison of CBT and supportive therapy regarding general symptoms of schizophrenia at post-treatment (data set #1), the comparison of family-based interventions to standard care regarding hospital admission between the 7th and 12th month of treatment (#11), and of CBT compared to TAU (#5) as well as compared to all control conditions (#6) in reducing positive symptoms. The Trim and Fill procedure suggested that there were missing studies in 15 of the 22 data sets (68.18%). As expected, all mean effect estimates were reduced if a correction was necessary. However, all of the 15 effect sizes were reduced non-significantly, as there was never a significant difference between the original and the corrected effect sizes, respectively. In addition, in most of the cases the general pattern of the efficacy of the interventions was not changed by taking a putative publication bias into account, as can be seen from the fact that both the original and the newly estimated corrected pooled effect sizes were significant in 12 of the 15 data sets in which studies had to be imputed. Both before and after correction, the efficacy of these therapeutic approaches was assessed as significantly larger than zero. Furthermore, for two data sets, the original and the corrected effect size estimates were both non-significant. In one data set, however, publication bias changed the pattern of efficacy, as the original *RR* did marginally fail to reach

significance, and the corrected *RR* turned significant. Therefore, the imputation of presumably missing studies resulted in a significant change of the effect size estimates only once. This result is most important: the impact of publication bias is only once severe and does otherwise not affect the efficacy of the treatments. In summary, there was thus very little evidence for a selective publication of positive results. The validity of the data sets that were tested and found to be not significantly affected by publication bias is enhanced. We can conclude that the minor tendency toward selective publication of positive results hardly changes the assessment of the efficacy for such interventions as CBT, family therapy, and psychoeducation.

Most of the meta-analyses provide evidence that the therapies investigated were of moderate to low efficacy. In 17 of the 22 data sets, the efficacy of psychotherapy for schizophrenia was significant and thus substantial. For these treatments, we can conclude that the efficacy is not the result of a publication bias. Moreover, in the data set in which the evaluation of efficacy was changed, the treatment condition was shown to be efficacious, as unexpectedly a study in favor of psychosocial family interventions was missing. Furthermore, psychotherapy with periods of unmedicated treatment was just as successful as psychological treatment combined with medication (#22). However, not all of the psychotherapeutic interventions for schizophrenia seem to be efficacious. Five of the data sets we included in the present study showed non-significant effect sizes. CBT was no more efficacious in reducing general symptoms than the control conditions (#1), as well as in reducing relapse (#2) or in reducing an early leaving of the study (#10). Furthermore, family interventions showed no benefit over standard care in an early leaving of the study, as well (#12, #13). These interventions can not be recommended for application in clinical practice.

It is important to note that asymmetry can result from many causes, such as undetected covariates, methodological inadequacies, or chance (Borenstein, Hedges, Higgins & Rothstein, 2009; Sterne et al., 2005). These other sources of asymmetry need to be considered when interpreting the significant results of the rank correlation and the regression method, for example poor methodological design of smaller studies or inadequate statistical analysis (Sterne et al., 2005). But whatever the reason for asymmetry, one always needs to question the mean effect estimate. If the efficacy in a single study depends on study size, it remains ambiguous how this intervention will work in routine care with individual patients (Sterne et al., 2005).

A limiting factor of the present study is that because many of the meta-analyses of schizophrenia did not provide the necessary raw data for a post-hoc assessment of publication bias, we were not able to include all published meta-analyses in our assessment. For this reason, guidelines for the conduction and publication of meta-analyses should be used in future psychotherapeutic research. The guidelines "Meta-analysis reporting standards" developed by the APA and integrated in the 6th edition of the APA Publication Manual (MARS; APA, 2008; 2009), and the "Preferred reporting items for systematic meta-analyses" (Prisma; Moher, Liberati, Tetzlaff & Altman, 2009) are already available and provide the necessary standards to conduct a meta-analysis. Both include recommendations for the reporting of the primary studies´ effect statistics and also routine application of methods to test for publication bias. We recommend that an assessment of small-study and publication bias should become the methodological standard. Sensitivity analyses for publication bias should routinely be included in every meta-analysis, rather than be left to follow-up analyses by other researchers (Borenstein et al., 2009; Rustenbach, 2003).

Publication bias and also the inclusion of unpublished studies were not considered in most of the meta-analyses. Only two meta-analyses of those we were able to include in the present study included unpublished studies (Jones et al., 2010; Pharoah et al., 2010). However, the unpublished studies were not integrated into the data sets we included from Pharoah et al. (2010), and only one unpublished study was included in the data set provided by Jones et al. (2010). Therefore, most of the data sets we assessed were based on published studies, and it is impossible to compare the results between data sets including only published and those including both published and unpublished studies. The large number of meta-analyses not including unpublished studies emphasizes the importance of drawing the attention of researchers to their inclusion and to the assessment of publication bias in meta-analyses. However, we can conclude that the low amount of publication bias found in the present study occurred even despite the integration of a mainly published body of research.

A method to detect publication bias that was applied in some of the meta-analyses is the failsafe N, initially developed by Rosenthal (1979), who emphasized the problem of publication bias and was the first to develop a statistical method for its assessment. Nevertheless, the failsafe N method by Rosenthal and also its variants developed since by other statisticians all suffer from many shortcomings and are of limited utility (Becker, 2005). For instance, the sample size of the omitted studies is not considered in the formula, equating studies with diverse sample sizes such as N=10 and N=10.000 (Begg & Berlin, 1988), and it is assumed that the missing studies have an average effect size of zero (Gleser & Olkin, 1996). In addition, there is no underlying statistical model and no statistically derived cut-off (Becker, 2005), and different formulas for the computation of the failsafe N lead to different estimates of the number of missing studies. The failsafe N therefore should be abandoned in favor of the statistical methods introduced above (Becker, 2005).

An approach based on selection models has also been proposed to detect publication bias (Hedges & Vevea, 2005). This very complex approach uses a weighted function of the p-values of each primary studies´ result, or of the primary studies´ effect size estimates and their associated standard errors to model the probability of publication. As with the Trim and Fill procedure, the number of missing studies and the effect of missing studies on the pooled effect size are estimated, but as selection models are difficult and highly computer intensive to run, they are not frequently used (Hedges & Vevea, 1996; 2005) and were not applied in the present study.

At present, cognitive behavioral therapy, family therapy, psychoeducation, IPT, cognitive remediation, social skills training, token systems, psychodynamic therapy, psychoanalysis, hypnosis and music therapy are the interventions for which sufficient numbers of studies exist that allowed for differential evaluation in meta-analyses with regard to different outcome measures. For other therapies, like humanistic therapy, interpersonal therapy or systemic therapy, no recommendations regarding their efficacy and application in practice can be given based on the present evidence. It thus seems either that such studies have not been conducted, or have not been published, which means that the efficacy of these therapies cannot be definitely assessed.

Solutions to the problem of a potential publication bias for therapies which were examined in a sufficient number of studies exist. First, the retrieval of unpublished studies by researchers conducting meta-analyses is important, but a search for unpublished studies is difficult and time-consuming. It may often be impossible to make sure that all unpublished studies in a field have been detected and included (Hopewell et al., 2005). The time and effort spent in searching them may not always be rewarded. Often the most that can be done is to

make indirect assessments of publication bias using meta-analyses (Moshagen & Musch, 2008), as was done in the present study. The solution that might be best suited to preventing publication bias is by implementing freely accessible registers, in which psychotherapy research studies are collected systematically and comprehensively at the time of their inception. Such registers would be capable of providing unbiased samples, because they are not selective in the inclusion of studies (Berlin & Ghersi, 2005; Dickersin, 2005; Rustenbach, 2003). Their major advantage is that studies included in such a register cannot go astray and can be included in future meta-analyses. Moreover, the researchers who conducted unpublished studies can be contacted to collect unpublished results. In fields of research in which registers have been implemented, the use of statistical post hoc methods to assess publication bias may eventually become obsolete.

It should be emphasized that publication bias does not provide evidence of malpractice on the side of individual scientists or editors. Rather, it can be explained by the intended filter function of significance. Significance of results is often wrongly viewed as a helpful mechanism of selection, because every year hundreds of studies are conducted in every field of research, which cannot be surveyed by any one researcher or practitioner. Actually, as defined by the sociologist Luhmann in his system theory, the constitutive distinction of science is “true/false” (Luhmann, 1990). Researchers strive to obtain true knowledge and results, and they should reject false results. This distinction has been partly replaced by the distinction “significant/non-significant”. Only significant results are regarded as interesting and meaningful, and hence worthy of publication, while non-significant results tend to seem disappointing and remain unpublished- a fate that should be reserved for wrong results. Various factors contribute to this problem, including the limited space available in scientific journals, the reluctance of researchers to invest work in submitting articles reporting negative

or null findings in anticipation of their likely rejection, as well as explicit or implicit editorial policies. However, this practice introduces systematic bias and threatens the validity of the published research in each field (Rustenbach, 2003).

The present analysis leads to the positive conclusion that there is at most a very minor tendency towards the selective reporting of positive results in research on the efficacy of therapeutic interventions addressing schizophrenia. Almost none of the meta-analyses we reanalyzed changed significantly when publication bias was taken into account. The only exception was a new pattern for the efficacy of psychosocial family interventions, which changed from non-significant to significant once the inverse publication bias towards the selective report of negative results we identified for this intervention was corrected for. In summary, not all treatments are efficacious for schizophrenia, but most of them are, and their efficacy does not seem to be the result of a reporting bias.

Acknowledgement

The authors would like to thank Helen-Rose Cleveland for language editing and Doris Heinrichs and Annett Schmitz for helpful technical assistance.

References

Meta-analyses are marked with an asterisk (*)

American Psychiatric Association (1987). Diagnostic and statistical manual of mental disorders (3rd ed. rev.). Washington, DC: American Psychiatric Association.

American Psychiatric Association (1994). Diagnostic and statistical manual of mental disorders DSM-IV (4th ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association (2000). Diagnostic and statistical manual of mental disorders: DSM-IV-TR (4th ed., rev.). Washington: DC: American Psychiatric Association.

American Psychological Association (2009). Publication manual of the American Psychological Association (6th ed.). Washington: DC American Psychiatric Association.

APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008). Reportings standards for research in psychology. Why do we need them? What might they be? Am. Psychol., 43, 839-851.

* Andrews, G. (1984). Treatment Outlines for the Management of Schizophrenia. The Quality Assurance Project. Aust. Nz. J. Psychiatry, 18, 19-38.

* Barbato, A. & D´Avanzo, B. (2000). Family interventions in schizophrenia and related disorders: a critical review of clinical trials. Acta Psychiatr. Scand., 102, 81-97.

Becker, B.J. (2005). Failsafe N or the file-drawer number. In: Rothstein, H.R., Sutton A.J., Borenstein M. (Eds.), Publication bias in meta-analysis. West Sussex: Wiley, pp. 111-126.

Begg, C.B. & Berlin, J.A. (1988). Publication bias: A problem in interpreting medical data (with discussion). J. Roy. Stat. Soc. A. Sta., 151, 419-463.

Begg, C.B. & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. Biometrics, 50, 1088-101.

* Benton, M.K. & Schroeder, H.E. (1990).Social Skills training with schizophrenics: a meta-analytic evaluation. J. Consult. Clin. Psych., 58, 741-747.

Berlin, J.A. & Ghersi, D. (2005). Preventing publication bias: Registries and prospective meta-analysis. In: Rothstein, H.R., Sutton A.J., Borenstein M. (Eds.), Publication bias in meta-analysis. West Sussex: Wiley, pp. 35-48.

* Bola, J.R. (2006). Medication-free Research in Early Episode Schizophrenia: Evidence of Long-Term Harm? Schizophr. Bull., 32, 288-296.

Borenstein, M., Hedges, L.V., Higgins, J.P.T. & Rothstein, H.R. (2009). Introduction to meta-analysis.West Sussex: Wiley & Sons.

Borenstein, M., Hedges, L.V., Higgins, J.P.T. & Rothstein, H.R. (2005). Comprehensive Meta-Analysis, Version 2.2. Englewood, NJ: Biostat.

* Corrigan, P.W. (1991). Social skills training in adult psychiatric populations: A meta-analysis. J. Behav. Ther. Exp. Psychiatry., 22, 203-210.

* Cuijpers, P. (1999). The effects of family interventions on relatives´ burden: a meta-analysis. J. Ment. Health, 8, 275-285.

Cuijpers, P., Smit, F., Bohlmeijer, E., Hollon, S.D. & Andersson, G. (2010). Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. Br. J. Psychiatry, 196, 173-178.

Degkwitz, R., Helmchen, H., Kockott, G. & Mombour, W., Hrsg. (1980). Diagnosenschlüssel und Glossar psychiatrischer Krankheiten: dt. Ausgabe d. internat. Klassifikation d. Krankheiten d. WHO, ICD, 9. Revision [International Classification of Diseases, Ninth Revision]. Chapter V. Springer, Berlin, Heidelberg, New York.

Dickersin, K. (2005). Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), Publication bias in meta-analysis. West Sussex: Wiley & Sons Ltd, pp. 11-34.

Duval, S. (2005). The Trim and Fill method. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), Publication bias in meta-analysis. West Sussex: Wiley & Sons Ltd, pp. 127-144.

Egger M., Davey Smith, G., Schneider, M. & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. Brit. Med. J., 315, 629-34.

Gleser, L.J. & Olkin, I. (1996). Models for estimating the number of unpublished studies. Stat. Med., 15, 2493-2507.

* Gould, R.A., Mueser, K.T., Bolton, E., Mays, V. & Goff, D. (2001).Cognitive therapy for psychosis in schizophrenia: an effect size analysis. Schizophr. Res., 48, 335-342.

* Hayes, R.L. & McGrath, J. (2009). Cognitive Rehabilitation for People with Schizophrenia and related conditions. Oxford, UK: The Cochrane Library, 1, 1-35.

Hedges, L.V. & Vevea, J. (1996). Estimating effect size under publication bias: Small sample properties and robustness of a random effects selection model. J. Educ. Behav. Stat., 21, 299-332.

Hedges, L.V. & Vevea, J. (2005). Selection method approaches. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), Publication bias in meta-analysis. West Sussex: Wiley & Sons Ltd, pp. 145-174.

* Hegarty, J., Baldessarini, R., Tohen, M. & Waternaux, C. (1994). One hundred years of schizophrenia: A meta-analysis of the outcome literature. Am. J. Psychiatry, 151, 1409-1416.

Hopewell, S., Clarke, M. & Mallett, S. (2005). Grey literature and systematic reviews. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), Publication bias in meta-analysis. West Sussex: Wiley & Sons Ltd, pp. 49-72.

Ioannidis, J.P.A. & Trikalinos, T.A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. Can. Med. Assoc. J., 176, 1091-1096.

* Izquierdeo de Santiago, A. & Khan, M. (2009). Hypnosis for Schizophrenia (Review). The Cochrane Library, Issue 4, 1-23.

* Jesus Mari, J. & Streiner, D. (1994). An overview of family interventions and relapse on schizophrenia: Meta-analysis of research findings. Psychol. Med., 24, 565-578.

* Jones, C., Cormac, I., Silveira da Mota Neto, J. I. & Campbell, C. (2010). Cognitive behaviour therapy for schizophrenia. The Cochrane Database of Systematic Reviews, *4*, 1-69.

*Krabbendam, L. & Aleman, A. (2003). Cognitive rehabilitation in schizophrenia: a quantitative analysis of controlled studies. Psychopharmacology, 169, 376-382.

* Kurtz, M.M. & Mueser, K.T. (2008). A Meta-Analysis of controlled Research on Social Skills Training for Schizophrenia. J. Consult. Clin. Psychol., 76, 491-504.

* Kurtz, M.M., Moberg, P.J., Gur, R.C. & Gur, R.E. (2001). Approaches to Cognitive Remediation of Neuropsychological Deficits in Schizophrenia: A Review and Meta-Analysis. Neuropsychol. Rev., 11, 197-210.

Light, R.J. & Pillemer, D.B. (1984). Summing up: The science of reviewing research. Cambridge, MA: Harvard University Press.

Light, R.J., Singer, J.D. & Willett, J.B. (1994). The visual presentation and interpretation of meta-analyses. In: Cooper, H.M. & Hedges, L.V. (Eds.), The Handbook of Research Synthesis. New York: Russell Sage Foundation.

* Lincoln, T.M., Wilhelm, K. & Nestoriuc, Y. (2007). Effectiveness of psychoeducation for relapse, symptoms, knowledge, adherence and functioning in psychotic disorders: A meta-analysis. Schizophr. Res., 96, 232-245.

* Lincoln, T.M., Suttner, C. & Nestoriuc, Y. (2008). Wirksamkeit kognitiver Interventionen für Schizophrenie. Eine Meta-Analyse. Psychologische Rundschau, 59, 217- 232.

Lipsey, M.W. & Wilson, D.B. (2001). Practical meta-analysis. Thousand Oakes: Sage Publications Ltd.

Luhmann, N. (1990). Die Wissenschaft der Gesellschaft [The science of society]. Frankfurt/M.: Suhrkamp.

* Lynch, D., Laws, K.R. & McKenna, P.J. (2010). Cognitive behavioural therapy for major psychiatric disorder: Does it really work? A meta-analytical review of well-controlled trials. Psychol. Med., 40, 9-24.

* Malmberg, L., Fenton, M. & Rathbone, J. (2010). Individual psychodynamic psychotherapy and psychoanalysis for schizophrenia and severe mental illness (Review). The Cochrane Library, Issue 3, 1-50.

* McGrath, J. & Hayes, R.L. (2009). Cognitive rehabilitation for people with schizophrenia and related conditions (review). The Cochrane Library, Issue 1, 1-33.

McGrath, J., Saha, S., Chant, D. & Welham, J. (2008). Schizophrenia: A concise overview of incidence, prevalence, and mortality. Epidemiol. Rev., 30, 67-76.

* McGurk, S.R., Twamley, E.W., Sitzer, D.I., McHugo, G.J. & Mueser, K.T. (2007). A meta-analysis of cognitive remediation in schizophrenia. Am. J. Psychiatry, 164, 1791-1802.

* McMonagle, T. & Sultana, A. (2009). Token economy for schizophrenia (Review). The Cochrane Library, Issue 1, 1-17.

Moher, D., Liberati, A., Tetzlaff, J. & Altman, D.G. (2009). Preferred reporting items for systematic reviews and meta-analysis: The PRISMA statement. Ann. Intern. Med., 151, 264 - 270.

* Mojtabai, R., Nicholson, R.A. & Carpenter, B.N. (1998). Role of psychosocial treatments in management of schizophrenia: A meta-analytic review of controlled outcome studies. Schizophr. Bull., 24, 569-587.

Moshagen, M. & Musch, J. (2008). Publication bias in studies on the efficacy of hypnosis as a therapeutic tool. Contemporary Hypnosis, 25, 94-99.

* Mueller, D.R., Roder, V. & Brenner, H.D. (2007). Effektivität des integrierten Psychologsichen Therapieprogramms für schizophren Erkrankte. Nervenarzt, 1, 62-73.

* Roder,V., Mueller, D.R., Mueser, K.T. & Brenner, H.D. (2006).Integrated Psychological Therapy (IPT) for Schizophrenia: Is it Effective? Schizophr. Bull., 32, 81-93.

* Mueller, D.R. & Roder, V. (2008). Empirical Evidence of Group Therapy addressing Social Perception in Schizophrenia. In: Teiford, J.B. (Ed.), Social Perception: 21th Century Issues and Challenges. Nova Science Publishers, Inc., pp. 51-80.

NICE (2003). Schizophrenia: Full National Clinical Guideline on Core Interventions in Primary and Secondary Care. Gaskell and the British Psychological Society: London.

NICE (2009). Schizophrenia: Core Interventions in the Treatment and Management of Schizophrenia in Primary and Secondary Care (update). National Institute for Clinical Excellence: London.

* Nose, M., Barbui, C., Gray, R. & Tansella, M. (2003). Clinical interventions for treatment non-adherence in psychosis: meta-analysis. Brit. J. Psychiatry, 183, 197-206.

* Pfammatter, M. (2005). Psychologische Therapieansätze bei Schizophren Erkrankten- Was wurde gemacht, was wird erreicht? (Doctoral dissertation). University of Bern, Switzerland.

* Pfammatter, M., Brenner, H.D. & Junghan, U.M. (2006). Efficacy of psychological therapy in schizophrenia: conclusions from meta-analyses. Schizophr. Bull., 32, 64-80.

* Pharoah, F., Mari, J., Rathbone, J. & Wong, W. (2010). Family intervention for schizophrenia (Review). The Cochrane Library, 12, 1-156.

* Pilling, S., Bebbington, P., Kuipers, E., Garety, P., Geddes, J., Orbach, G. & Morgan, C. (2002). Psychological Treatments in Schizophrenia: I. Meta-Analysis of Family Intervention and Cognitive Behaviour Therapy. Psychol. Med., 32, 763-782

*Pilling, S., Bebbington, P., Kuipers, E., Garety, P., Geddes, J., Orbach, G. & Morgan, C. (2002). Psychological treatments in schizophrenia: II. Meta-analysis of randomized controlled trails of social skills training and cognitive remediation. Psychol. Med., 32, 783-791.

* Pitschel-Walz, G., Leucht, S., Bäuml, J., Kissling, W. & Engel, R. (2001). The effect of family interventions on relapse and rehospitalization in schizophrenia - A meta-analysis. Schizophr. Bull., 27, 73-92.

* Pitschel-Walz, G. (1997). Die Einbeziehung von Angehörigen in die Behandlung schizophrener Patienten und ihr Einfluss auf den Krankheitsverlauf. Europäische Hochschulschriften, Reihe 6, Vol. 581. Frankfurt a.M.: Peter Lang Verlag.

* Pitschel-Walz, G. & Engel, R. (1997). Psychoedukation in der Schizophreniebehandlung; Einfluss auf den Krankheitsverlauf. Europäische Hochschulschriften, 23. Frankfurt a.M.: Peter Lang Verlag.

* Rector, N.A. & Beck, A.T. (2001). Cognitive Behavioral Therapy for Schizophrenia: An Empirical Review. J. Nerv. Ment. Dis., 189, 278-287.

Rothstein, H.R., Sutton, A.J. & Borenstein, M. (2005). Publication bias in meta-analysis. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), Publication bias in meta-analysis. West Sussex: Wiley & Sons Ltd, pp. 1-8.

Rosenthal, R. (1979). The „file drawer problem“ and tolerance for null results. Psychol. Bull., 86, 638-641.

Rustenbach, S.J. (2003). Metaanalyse. Eine anwendungsorientierte Einführung [Meta-analysis. An applicationDie Einbeziehung der Angehörigen in die Behandlung schizophrener Patienten und ihr -oriented introduction]. Verlag Hans Huber, Bern.

* Sepehry, A.A., Potvin, S., Elie, R. & Stip, E. (2007). Selective Serotonin Reuptake Inhibitor (SSRI) add-on therapy for the negative symptoms of schizophrenia: A meta-analysis. J. Clin. Psychiatry, 68, 604-610.

* Silverman, M.J. (2003). The influence of music on the symptoms of psychosis: a meta-analysis. J. Music Ther., 40, 27-40.

Slade, M. & Priebe, S. (2001). Are randomized controlled trials the only gold that glitters? BJP 2001, 179, 286-287.

Sterne, J.A.C., Becker, B.J. & Egger, M. (2005). The funnel plot. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), Publication bias in meta-analysis. West Sussex: Wiley & Sons Ltd, pp. 75-98.

Sterne, J.A. & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), Publication bias in meta-analysis. West Sussex: Wiley & Sons Ltd, pp. 99-110.

Sterne, J.A., Gavaghan, D. & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. J. Clin. Epidemiol., 53, 1119-1129.

* Suslow, T., Schonauer, K. & Arolt, V. (2001). Attention training in the cognitive rehabilitation of schizophrenic patients: A review of efficacy studies. Acta Psychiatr. Scand., 103, 15-23.

Sutton, A.J., Duval, S.J., Tweedie, R.L., Abrams, K.R. & Jones, D.R. (2000). Empirical assessment of effect of publication bias on meta-analyses. Brit. Med. J., 320, 1574-1577.

* Tarrier, N. (2005). Cognitive behaviour therapy for schizophrenia- a review of development, evidence and implementation. Psychother. Psychosom., 74, 136-144.

* Tarrier, N. & Wykes, T. (2004). Is there evidence that cognitive behaviour therapy is an effective treatment for schizophrenia? A cautious or cautionary tale? Beh. Res. Ther., 42, 1377-1401.

Taylor, S. & Tweedie, R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. J. Am. Stat. Assoc., 95, 89-98.

Taylor, S. & Tweedie, R. (2000b). Trim and Fill: A simple funnel-plot based method of testing and adjusting for publication bias in meta-analysis. Biometrics, 56, 455-63.

Terrin, N., Schmid, C.H., Lau, J. & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. Stat. Med., 22, 2113–2126.

* Twamley, E.W., Jeste, D.V. & Bellack, A.S. (2003). A review of cognitive training in schizophrenia. Schizophr. Bull., 29, 359-382.

* Villeneuve, K., Potvin, S., Lesage, A. & Nicole, L. (2010). Meta-analysis of rates of drop-out from psychosocial treatment among persons with schizophrenia spectrum disorder. Schizophr. Res., 121, 266-270.

World Health Organisation (1992). The ICD-10 classification of mental and behavioral disorders: Clinical descriptions and diagnostic guidelines. Geneva: Author.

* Wunderlich, U., Wiedemann, G. & Buchkremer, G. (1996). Sind psychosoziale Interventionen bei schizophrenen Patienten wirksam? Eine Meta-Analyse. Verhaltenstherapie, *6*, 4-13.

* Wykes, T., Steel, C., Everitt, B. & Tarrier, N. (2008). Cognitive behavior therapy for schizophrenia: Effect sizes, clinical models, and methodological rigor. Schizophr. Bull., 34, 523-537.

* Zimmermann, G., Favrod, J., Trieu, V.H. & Pomini, V. (2005). The effect of cognitive behavioral treatment on the positive symptoms of schizophrenia spectrum disorders: A meta-analysis. Schizophr. Res., 77, 1-9.

Table 1: Results of the Trim and Fill procedure, Begg and Mazumdar's (1994) rank-correlation test, and Egger et al.´s (1997) regression analysis.

Data set number	Author	Intervention/ dependent measure/ time of measurement	Original effect size (95% CI)	Replicated effect size (95% CI)	Corrected effect size [+] (95% CI)	Number of studies included	Number of studies missing [++]	Begg (Tau) and Egger (β_0)
1	Lynch et al. (2010)	CBT vs. control/general symptoms/post	g = -0.08 (-0.24; 0.08) (FE)	-0.08 (-0.24; 0.08)		9	0	Tau = -.61*† β_0 = -3.35*†
2	Lynch et al. (2010)	CBT vs. control/relapse/follow-up	OR = 1.17 (0.88; 1.55) (FE)	1.17 (0.88; 1.55)	0.96 (0.75; 1.23)	8	1	Tau = -.07 β_0 = -.74
3	Lincoln et al. (2008)	CBT vs. TAU/general symptoms/ post	g = 0.25[a] (0.14; 0.36) (FE)	0.26[a] (0.14; 0.37)	0.24[a] (0.13; 0.36)	9	1	Tau = .00 β_0 = .68
4	Wykes et al. (2008)	CBT vs. control/low quality/ positive symptoms/post	d = 0.49[a] (0.31; 0.66) (RE)	0.45[a] (0.31; 0.60)	0.42[a] (0.25; 0.58)	20	2	Tau = .12 β_0 = .6
5	Zimmermann et al. (2005)	CBT vs. TAU/positive symptoms/ post	g = 0.30[a] (0.15; 0.46) (FE)	0.30[a] (0.15; 0.46)	0.21[a] (0.07; 0.35)	8	3	Tau = .21 β_0 = 1.83*
6	Zimmermann et al. (2005)	CBT vs. control/positive symptoms/post	g = 0.35[a] (0.23; 0.47) (FE)	0.35[a] (0.23; 0.47)	0.26[a] (0.15; 0.38)	14	5	Tau = .1 β_0 = 1.45*†
7	Zimmermann et al. (2005)	CBT vs. control/positive symptoms/chronic samples/post	g = 0.27[a] (0.11; 0.42) (FE)	0.29[a] (0.15; 0.43)	0.20[a] (0.07; 0.32)	10	4	Tau = -.04 β_0 = 1.22
8	Zimmermann et al. (2005)	CBT vs. TAU/positive symptoms/ chronic samples/post	g = 0.26[a] (0.09; 0.43) (FE)	0.26[a] (0.09; 0.43)	0.17[a] (0.01; 0.33)	6	2	Tau = .07 β_0 = 1.7
9	Zimmermann et al. (2005)	CBT vs. nonspecific treatment/ positive symptoms/post	g = 0.42[a] (0.21; 0.64) (FE)	0.43[a] (0.21; 0.64)		7	0	Tau = .05 β_0 = .21
10	Jones et al. (2010)	CBT vs. standard care/leaving study early/follow-up	RR = 0.80 (0.58; 1.1) (RE)	0.80 (0.58; 1.11)		7	0	Tau = -.14 β_0 = -.41

Data set number	Author	Intervention/ dependent measure/ time of measurement	Original effect size (95% CI)	Replicated effect size (95% CI)	Corrected effect size [+] (95% CI)	Number of studies included	Number of studies missing [++]	Begg (Tau) and Egger (β_0)
11	Pharoah et al. (2010)	Family interventions vs. standard care/hospital admission/7-12 month	RR = 0.78[a] (0.63; 0.98) (FE)	0.79 (0.60; 1.04)		9	0	Tau = -.67 *† β_0 = -3.44 *†
12	Pharoah et al. (2010)	Family interventions vs. standard care/leaving study early/7-12 month	RR = 0.74 (0.53; 1.03) (FE)	0.77 (0.55; 1.09)	0.76 (0.54; 1.07)	10	1	Tau = .14 β_0 = .58
13	Pharoah et al. (2010)	Family interventions vs. standard care/leaving study early/13-24 month	RR = 0.74 (0.55; 1) (FE)	0.75 (0.55; 1)	0.73[a] (0.55; 0.98)	10	1	Tau = .2 β_0 = .35
14	Pitschel-Walz (1997)	FT & PE vs. standard care/ relapse/post & follow-up combined/ schizophrenia	r = 0.21 (n.s.)[b]	0.22[a] (0.15; 0.29) (FE)	0.20[a] (0.13; 0.27)	9	2	Tau = .11 β_0 = .15
15	Pitschel-Walz (1997)	FT & PE vs. standard care/re-lapse/post & follow-up combined	r = 0.19[a] (0.13; 0.24)[b]	0.19[a] (0.13; 0.25) (FE)	0.18[a] (0.12; 0.24)	12	1	Tau = .03 β_0 = .08
16	Pitschel-Walz (1997)	FT & PE vs. standard care/ relapse/post/short-term interventions	r = 0.14 (n.s.)[b]	0.15[a] (0.06; 0.23) (FE)	0.13[a] (0.05; 0.22)	7	1	Tau = -.05 β_0 = .16
17	Pitschel-Walz (1997)	FT & PE vs. standard care/ relapse/post	r = 0.18 (n.s.)[b]	0.20[a] (0.14; 0.26) (FE)	0.18[a] (0.12; 0.24)	10	2	Tau = .18 β_0 = .36
18	Pitschel-Walz & Engel (1997)	FT & PE vs. standard care/ relapse/post & follow-up combined	r = 0.20[a] (0.15; 0.25)[b]	0.20[a] (0.15; 0.25) (FE)		17	0	Tau = -.08 β_0 = .12

Data set number	Author	Intervention/ dependent measure/ time of measurement	Original effect size (95% CI)	Replicated effect size (95% CI)	Corrected effect size [+] (95% CI)	Number of studies included	Number of studies missing [++]	Begg (Tau) and Egger (β_0)
19	Pitschel-Walz (1997)	FT & PE vs. standard care/general symptoms/post & follow-up combined	r = 0.20 (n.s.)[b]	0.23[a] (0.12; 0.33) (FE)	0.18[a] (0.08; 0.28)	6	2	Tau = .11 β_0 = 1.24
20	Pitschel-Walz (1997)	PE vs. standard care/relapse/post & follow-up combined	r = 0.14 (n.s.)[b]	0.14[a] (0.05; 0.23) (FE)		6	0	Tau = .33 β_0 = .47
21	Pitschel-Walz et. al. (2001)	FT & PE vs. standard care/re-lapse/post & follow-up combined	r = 0.20[a] (0.14; 0.27)[b]	0.19[a] (0.12; 0.26) (FE)	0.16[a] (0.09; 0.22)	14	3	Tau = .03 β_0 = .89
22	Bola (2006)	Unmedicated vs. medication/ global assessment/follow-up	r = -0.09 (-0.27; 0.09) (FE)	-0.08 (-0.16; 0.01)		6	0	Tau = .2 β_0 = .29

Note. CBT = Cognitive behavioral therapy; CI = confidence interval; d = Cohen´s d; FE = fixed effects model; g = Hedges´ g; n.s. = confidence limits not specified; OR = odds ratio; PT = psychotherapy; r = correlation; RE = random effects model; RR = relative risk; [+] = for publication bias; [++] =according to Trim and Fill

[a] significant effect size with a 95% confidence interval not including zero (for RR and OR: confidence interval not including 1)

[b] the integration model was not specified in the paper

* $p < .05$, one-sided

† $p < .05$, two-sided

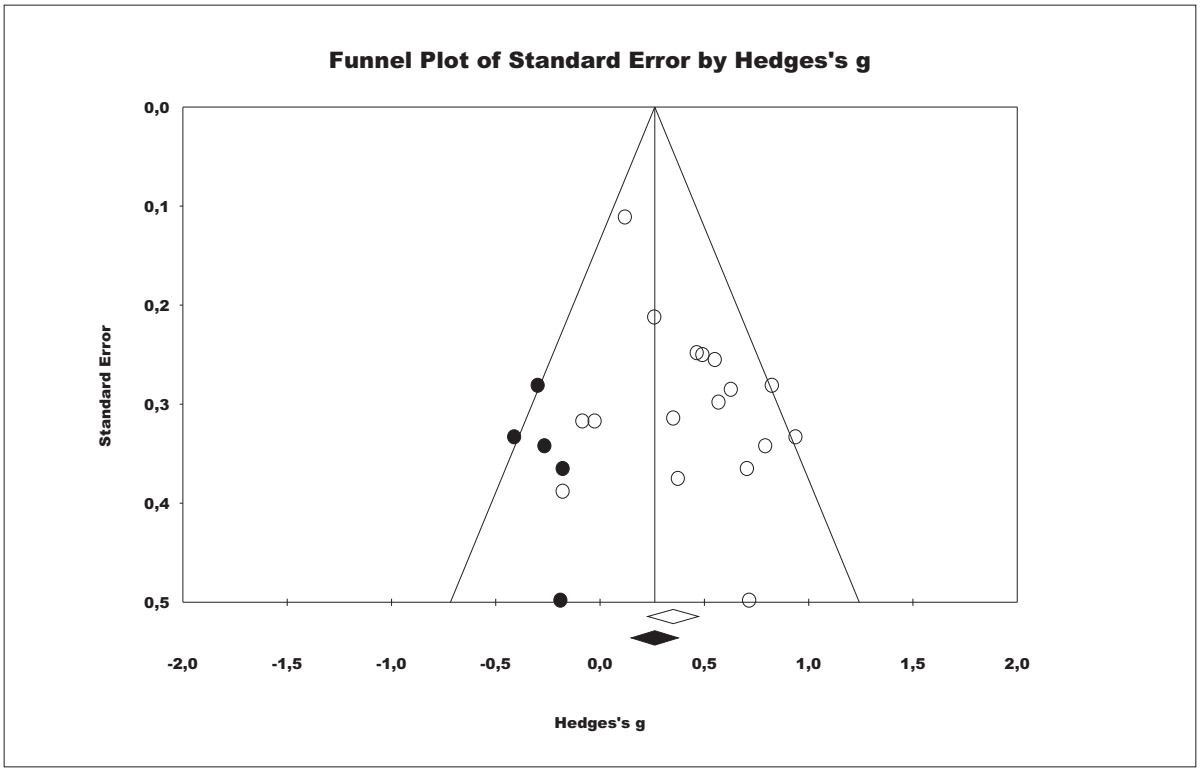

Figure 1.

An illustration of the Trim and Fill method showing the funnel plot for the reduction of reduction of positive symptoms through CBT at post-test (data set #6; Zimmermann et al., 2005). Dark circles: the 5 studies identified as missing.

15.10.11

Dear Mrs. Helen Niemeyer,

Your submission entitled "Publication Bias in Meta-Analyses of the Efficacy of Psychotherapeutic Interventions for Depression" has been received by Clinical Psychology Review

Your manuscript will be given a reference number once an Editor has been assigned.

Thank you for submitting your work to this journal.

Kind regards,

Elsevier Editorial System
Clinical Psychology Review

Publication Bias in Meta-Analyses of the Efficacy of Psychotherapeutic Interventions for Depression

Helen Niemeyer, Jochen Musch, Reinhard Pietrowsky

Heinrich-Heine-University, Department of Experimental Psychology, Universitätsstraße 1, 40225 Düsseldorf, Germany.

Corresponding author:

Helen Niemeyer, Heinrich-Heine-University, Department of Experimental Psychology, Universitätsstraße 1, 40225 Düsseldorf, Germany; Tel: ++49 211 81-12272; FAX: ++49 211 81-14261; Email: helen.niemeyer@hhu.de

None of the authors has any conflict of interest.

Abstract

Objective: The aim of this study was to assess whether systematic reviews investigating psychotherapeutic interventions for depression are affected by publication bias.

Method: We applied Begg and Mazumdar´s adjusted rank correlation test, Egger´s regression analysis, and the Trim and Fill procedure to assess the presence and magnitude of publication bias in all systematic reviews published up to September 2010.

Results: 31 data sets reported in 19 meta-analyses fulfilled our inclusion criteria. Significant bias was detected in 5 (16.13%; rank correlation test) and 6 (19.35%; Egger's regression analysis) of these data sets. Applying the Trim and Fill procedure to amend presumably missing studies rarely changed the assessment of the efficacy of therapeutic interventions with two exceptions. Psychotherapy was no longer found to be significantly more efficacious than pharmacotherapy when publication bias was taken into account. Moreover, after correcting for publication bias, there was no longer evidence that depressed patients without comorbid personality disorder (PD) profited more from psychotherapy than patients without comorbid PD.

Discussion: The results suggest that taken together, psychotherapy research for depression is only marginally affected by the selective reporting of positive outcomes. With two notable exceptions, correcting for publication bias did not change the evaluation of the efficacy of psychotherapeutic interventions.

Keywords

Depression, publication bias, meta-analysis, psychotherapy research

Publication Bias in Meta-Analyses of the Efficacy of Psychotherapeutic Interventions for Depression

Epidemiological studies report that depression affects 5 - 10% of the general population worldwide (Singleton, Bumpstead, O'Brien, Lee & Meltzer, 2001; Demyttenaere, Bruffaerts, Posada-Villa et al., 2004). Prevalence rates are high even in children (2.5%) and adolescents (8.3%; Birmaher et al., 1996). As depression impairs function and is linked to an increased risk of suicide (Hirschfeld et al., 1997), efficacious treatments are essential to reduce the suffering. Several evidence-based treatments with high efficacy do exist, such as cognitive behavioural therapy (CBT) or interpersonal therapy (IPT), and their efficacy for the treatment of depression has been investigated in a large number of empirical studies. Similarly, prevention programs have also been evaluated thoroughly. To obtain an overview over the existing wealth of studies and to achieve results with increased inferential power and validity, meta-analyses have been used to synthesize the results of these primary studies (Rustenbach, 2003). Therefore, evidence from meta-analyses is regarded as being the strongest on the widely used hierarchy of evidence (Slade & Priebe, 2001), and the evidence base for CBT and other psychotherapeutic interventions for depression is grounded primarily on meta-analytic results.

The state-of-the-art procedures used in evidence-based psychotherapy derive in the main from published rather than unpublished research (Gilbody & Song, 2000). Published results are most likely to be retrieved in literature searches, and therefore more often included in meta-analyses (Jackson, 2006). However, the validity of the effect size estimates for therapeutic interventions can be seriously impaired by publication bias if meta-analyses are mostly based on published studies, while unpublished studies draw a different picture (Rothstein, Sutton & Borenstein, 2005; Jackson, 2006). Publication bias is a systematic bias

characterized by the selective publication of studies with positive results, as opposed to studies with null or negative results (Hopewell, Clarke & Mallett, 2005).

To obtain more valid results it is important to address the risk of publication bias. If the sample of included studies is flawed, the validity of the results of a meta-analysis may be seriously compromised, even if the analysis is of otherwise high quality. If studies are not missing at random, but are systematically left out due to the non-significance or negativity of their results, publication bias may lead to an overestimation of the efficacy of therapeutic interventions, and the additional consideration of unpublished results might reduce the overall effect size estimate and have severe practical implications if it changes the assessment of the efficacy of a therapeutic intervention (Rustenbach, 2003).

Therapeutic interventions in clinical practice should be evidence-based, that is clinical decision making should be based on "best practice" (Slade & Priebe, 2001). If publication bias is not taken into account, psychotherapists may be led to using interventions seemingly based on apparent empirical evidence that however are only erroneously supported by the results of meta-analyses (Berlin & Ghersi, 2005). The application of therapies falsely viewed as being helpful due to overestimated effect sizes can lead to inefficacious treatment of depression. Severe side effects of therapies may exist and yet remain unknown if studies reporting such side effects are not being published (Rothstein et al., 2005). Treatments that are less efficacious than assumed also result in unnecessarily high costs for society and the health care system. Publication bias can also result in harm for patients, wasted resources, and a misguidance of future research (Blumle, Antes, Schumacher, Just & Von Elm, 2008; Von Elm et al., 2008). An ethical obligation to publish the results of a study once it is finished exists both with respect to the patients having taken part in that study, and with respect to other patients suffering from depression (Chalmers, 1990; Pearn, 1995). The non-publication of

research findings violates this ethical obligation, and poses a serious threat to the possibility of fully informed health care decisions (McGauran et al., 2010).

Sterling (1959) conducted the first analyses of publication bias and found that in psychology, most of the published studies reported significant results. This finding was subsequently widely confirmed for the fields of social science (Glass, McGaw & Smith, 1981; Smith, 1980a) and medicine (Moscati, Jehle, Ellis, Fiorello & Landi, 1994; Vickers, Gloyal, Harland & Rees, 1998). As publication bias can occur in all fields of research, approaches to detect and reduce publication bias should always be applied. They encompass the search for unpublished studies when conducting a meta-analysis, the registration of all studies at the time of their inception, and a search for publication bias in meta-analyses using formal statistical methods (Berlin & Ghersi, 2005; Moshagen & Musch, 2008; Rustenbach, 2003).

Quantitative and hence objective methods to detect publication bias are available and include Begg and Mazumdar´s rank correlation method (1994), Egger´s regression analysis (Egger, Davey Smith, Schneider & Minder, 1997), and the Trim and Fill procedure (Taylor & Tweedie, 2000a, 2000b). These methods, all based on the funnel plot, test for asymmetry in the dispersion of the primary studies´ effect sizes around the mean effect estimate. If bias is detected, the Trim and Fill procedure additionally computes the number of presumably missing studies and a revised effect size estimate which is corrected for bias (Duval, 2005).

The key assumption of all methods based on the funnel plot is that observed asymmetries are due to publication bias. It is important to note, however, that the cause of an asymmetry cannot by determined from the asymmetry itself (Ioannidis & Trikalinos, 2007; Sterne, Gavaghan & Egger, 2000; Terrin, Schmid, Lau & Olkin, 2003). Rather, asymmetry can result from many causes, including heterogeneous effects, undetected covariates,

methodological inadequacies or chance (Sterne, Becker & Egger, 2005). If there is significant between-study heterogeneity in a data set, researchers usually try to find the reasons for this heterogeneity by applying techniques such as subgroup analysis or meta-regression (Borenstein, Hedges, Higgins & Rothstein, 2009). It is important to note, however, that statistical methods based on a homogeneity assumption are inappropriate and lead to false alarms if evidence for heterogeneity is found (Ioannidis & Trikalinos, 2007; Sterne et al., 2000; Terrin et al., 2003). In a reanalysis of existing data sets, Ioannidis and Trikalinos (2007) found that in most meta-analyses the homogeneity requirement was in fact not met. Moreover, they found that if publication bias was detected, heterogeneity frequently offered an alternative explanation for the observed asymmetry.

In order to detect publication bias in studies investigating the efficacy of psychotherapeutic interventions for depression, Cuijpers, Smit, Bohlmeijer, Hollon and Andersson (2010) recently conducted a meta-analytic study on publication bias. They selected all randomized controlled studies in which the efficacy of different psychotherapeutic treatments for depression was investigated from a database of published studies, www.evidencebasedpsychotherapies.org. Overall, they included 117 primary studies and calculated pooled mean effect size estimates for CBT, as well as for a sample of all other psychotherapeutic interventions taken together, and applied Begg and Mazumdar´s rank correlation method (Begg & Mazumdar, 1994), Egger´s regression analysis (Egger et al., 1997), and the Trim and Fill procedure (Taylor & Tweedie, 2000a, 2000b) for a statistical assessment of publication bias. A substantial degree of publication bias was detected, and Cuijpers, Smit, et al. therefore concluded that the efficacy of both CBT and other types of psychotherapy was most likely overestimated. However, heterogeneity was highly significant in 40 out of the 43 data sets investigated by Cuijpers, Smit, et al. (2010), offering an

alternative explanation for their findings. Moreover, there is a large number of studies investigating the efficacy of depression treatments that was not included in the meta-analysis of Cuijpers, Smit, et al. (2010). The present study was therefore conducted to provide a more comprehensive assessment of publication bias in the reporting of studies investigating the efficacy of psychotherapeutic interventions for depression. We aimed at re-assessing all existing meta-analyses of therapeutic interventions for the potential presence of publication bias. The vast majority of these meta-analyses either did not use formal statistical methods to detect the presence of publication bias, or did only employ one of the several methods that are available for this purpose. To avoid the false positive identification of publication bias in meta-analyses in which in fact heterogeneous studies led to asymmetry in the funnel plot (Ioannidis & Trikalinos, 2007), we also decided to set the homogeneity of the data as inclusion criterion (Ioannidis, 2005; Terrin et al., 2003).

A literature search up to September 2010, aimed at identifying all meta-analyses investigating the efficacy of psychotherapeutic interventions for depression (not combined with other disorders) and the prevention of depression, resulted in 85 meta-analyses. In these meta-analyses the following types of interventions were investigated: CBT, behavioral activation, problem-solving therapy, the “Coping with depression”-course, IPT, relaxation, psychoeducation, a combination of several types of psychotherapy (including some of the aforementioned as well as nondirective therapy, for example), a combination of psychotherapy and pharmacotherapy, couple therapy, group therapy, internet-based therapy, short-term psychodynamic therapy, positive psychology, reminiscence therapy, case management in primary health care and also preventive interventions against depression. Some of these interventions were investigated in particular samples, such as adolescents, the elderly, in- and outpatient groups, women after birth, or patients suffering from somatic

diseases and comorbid depression (all of these meta-analyses are marked with an asterisk in the list of references). Most of these meta-analyses provided evidence that the therapies investigated are of moderate to large efficacy. However, the validity of meta-analyses that are limited to the inclusion of published studies is severely threatened by a systematic bias (Rustenbach, 2003), and only in 16 of the 85 meta-analyses (18.82%), unpublished studies were included (Bohlmeijer, Smit & Cuijpers, 2003; Bortolotti, Menchetti, Bellini, Montaguti & Berardi, 2008; Cuijpers, 1998b; Cuijpers, Li et al., 2010; Cuijpers, Van Straten & Smit, 2006; Driessen et al., 2010; Ekers et al., 2008; Gloaguen, Cottraux, Cucherat & Blackburn, 1998; Gregory, Schwer Canning, Lee & Wise, 2004; Kühner, 2003; Mazzuchelli et al., 2009; Michael & Crowley, 2002; Neumeyer-Gromen, Lampert, Stark & Kallischnigg, 2004; Sin & Lyubomirsky, 2009; Spek et al., 2007; Weisz, McCarty & Valeri, 2006). Without the use of formal statistical methods to detect the presence of publication bias, it is impossible to tell whether the selective publication of positive results has distorted the effect size estimates reported in these meta-analyses. Unfortunately, most of the meta-analyses did not employ formal statistical methods, or employed only one of several statistical methods that are useful for this purpose. In particular, the Trim and Fill procedure (Taylor & Tweedie, 2000a, 2000b) had been conducted in only 12 (14.12%) of these 85 meta-analyses (Andersson & Cuijpers, 2009; Cuijpers, Brännmark & Van Straten, 2008; Cuijpers, Li, Hofmann & Andersson, 2010; Cuijpers, Smit, et al., 2010; Cuijpers, Smit & Van Straten, 2007; Cuijpers, Van Straten, Andersson & Van Oppen, 2008; Cuijpers, Van Straten, et al., 2010; Cuijpers, Van Straten, Hollon & Andersson, 2010; Cuijpers, Van Straten, Smit, Mihalopoulos & Beekman, 2008; Driessen et al., 2010; Mazzuchelli, Kane & Rees, 2009; Watanabe, Hunot, Omori, Churchill & Furukawa, 2007). Moreover, in only 3 of the 85 meta-analyses (3.53%), Begg and Mazumdar´s rank correlation test (1994) had been conducted (Cuijpers, Smit, et al., 2010;

Ekers, Richards & Gilbody, 2008; Himelhoch, Medoff & Oyeniyi, 2007), and only 6 meta-analyses (7.06%) applied Egger et al.´s (1997) regression analysis (Cuijpers, Smit, et al., 2010; Ekers et al., 2008; Gensichen et al., 2006; Himelhoch et al., 2007; Mazzuchelli et al., 2009; Watanabe et al., 2007). These results indicate a wide-spread lack of statistical control for publication bias; existing publication bias may therefore well have gone undetected in previous meta-analyses, and the efficacy of the interventions therefore may have been overestimated. For this reason, we conducted a comprehensive analysis of publication bias in meta-analyses of the efficacy of therapeutic interventions by employing the three most frequently recommended formal procedures to detect publication bias. However, unlike previous meta-analyses, we limited our sample to homogeneous data sets to avoid false positives that have been argued to occur if these formal procedures are used inappropriately (Ioannidis & Trikalinos, 2007). Our analysis therefore constitutes a more conservative - and arguably more valid - test for the presence of publication bias than has been published before.

Methods

Data sources

To comprehensively identify all relevant meta-analyses on psychotherapeutic interventions and prevention programs for depression, we conducted a literature search in the databases PsycInfo and Psyndex, as well as in reference lists of articles and book chapters, following the search strategies recommended by Lipsey and Wilson (2001). Articles were retrieved for further assessment if the title or abstract suggested that a meta-analysis concerning psychotherapy or prevention methods for depression had been conducted. All available meta-analyses up to September 2010 were included. The keywords “meta-analysis” or “systematic review” combined with ”depression”, “major depression”, ”depressive

disorder" and "dysthymia" were used in all variations. The search was restricted to reviews and meta-analyses reported in either English or German language.

Study selection and data extraction

All articles, including those for which the abstract provided insufficient information, were thoroughly examined in order not to miss any relevant meta-analyses. Eligible meta-analyses had to involve major depression or dysthymic disorder according to the diagnostic criteria of either DSM-III (APA, 1980), DSM-III-R (APA, 1987), DSM-IV (APA, 1994), DSM-IV-TR (APA, 2000), ICD-9 (Degkwitz, Helmchen, Kockott & Mombour, 1980) or ICD-10 (World Health Organization, 1992). Dysphoric samples with high scores on self-rating scales for depressive symptomatology were also included. Eligible meta-analyses for the prevention of depression or relapses did not have to include samples with a current episode of major depression, but had to measure depressive symptoms before and after the application of the intervention. The intervention also had to aim at preventing the development of a depressive disorder or a relapse. At least one psychotherapeutic or preventive intervention had to be involved to make a meta-analysis eligible for inclusion. Interventions were not restricted to a specific school of psychotherapy. Reviews in which primary studies with various disorders were pooled for the mean effect size calculations, and which thus were not restricted to the treatment of depression, were excluded. Finally, all meta-analyses were excluded that targeted the effectiveness of pharmacological treatment or electroconvulsive treatment.

A pooled effect size estimate, the primary studies´ effect sizes and the measures of their precision are necessary to assess publication bias. All arms of studies which fulfilled the following inclusion criteria were therefore included: 1) all effect sizes of the primary studies,

or raw data for their calculation, were given; 2) a pooled effect size estimate was reported and could be confirmed using the original primary studies´ effect sizes; 3) a measure of precision (confidence interval, standard deviation, standard error or variance) was available for the raw data or primary studies´ effect sizes. If the necessary data were not reported, an attempt was made to obtain them from the authors. If the original study reported a correlation coefficient but no measure of precision, the sample size was used to calculate this measure. Additional inclusion criteria were that 4) data sets were homogeneous, to avoid a violation of the respective assumption underlying all statistical methods to assess funnel plot asymmetry; and 5) at least six studies had to be included, because the detection of publication bias is unreliable and of low power (Sterne et al., 2005) if less than six studies are analyzed (Egger et al., 1997). Finally, 6) we only included studies for which an effect size estimate corrected for publication bias had not yet been calculated using the Trim and Fill procedure.

Assessment of publication bias

As mentioned above, several procedures have been proposed for the assessment of publication bias. The most widely used methods are based on the funnel plot (Light & Pillemer, 1984). Funnel plots chart the primary studies´ effect sizes (the treatment estimates) against a measure of their precision. If no bias is present, the funnel plot is symmetric about the mean effect size, with equal dispersion around the mean at any level of precision. The effect size estimates at the bottom of the graph from smaller and thus less precise studies will vary more than those from larger studies, and will therefore scatter more widely at the base of the plot. The spread narrows towards the top of the funnel, due to the higher precision of larger studies. Under the premise that smaller studies with null or negative effects are non-randomly excluded from publication, the funnel plot becomes asymmetrical at its base (Gilbody & Song, 2000; Sterne et al., 2005). Since a visual assessment of asymmetry in the

dispersion is subjective and thus unreliable (Terrin et al., 2003), three objective methods have been developed, all of which were applied in the present study: Begg and Mazumdar´s rank correlation method (1994), Egger´s regression analysis (Egger et al., 1997), and the Trim and Fill procedure (Taylor & Tweedie, 2000a, 2000b).

Begg and Mazumdar‘s rank correlation method and Egger‘s regression analysis both investigate whether a correlation between the variances and the effect sizes of the primary studies is present. In case of publication bias small studies with higher variances are more likely to be published if they show large effect sizes. This leads to a positive association between variance and effect size, which is also called small study bias. The non-parametric adjusted rank correlation method detects small study effects by assigning ranks to the standardized effect sizes and their variances and examining the rank correlation between them. There is no relationship between the effect sizes and their variances in the absence of bias. The significance of correlation between the ranks is tested using Kendall´s Tau (Sterne & Egger, 2005). In the parametric linear regression method the standard normal deviate - effect size divided by its standard error - is regressed against its precision, given by the inverse of its standard error. Magnitude and direction of the effect are indicated by the slope, while the intercept provides a measure of the degree of asymmetry, in which case the intercept does not run through the origin (Sterne & Egger, 2005). We conducted both one- and two-sided significance tests with the type I error level set at .05. This decision was made because while under the typical scenario of publication bias, a one-tailed test is usually conducted (Cuijpers, Smit, et al., 2010), in a scenario where larger effects in larger studies may also be observed, a two-tailed test is more appropriate (Sterne & Egger, 2005).

Neither Begg´s nor Egger´s method adjusts for publication bias. Beyond the mere existence of bias, the non-parametric Trim and Fill procedure (Taylor & Tweedie, 2000a,

2000b) provides a corrected effect size estimate taking the bias into account. Rather than investigating the correlation between the effect size and its variance, the mean effect estimate is used as a fixed point to test which studies with positive effect sizes have no mirror image counterparts with negative effect sizes. In an algorithm the rightmost studies in the funnel plot with the largest effect sizes, which do not have counterparts on the left side of the mean effect size estimate are trimmed off, and the mean effect is re-estimated. Then, the studies and their missing counterparts are imputed. After a few iterations the trimmed and filled funnel plot has usually become symmetric and the procedure stops, outputting a new bias-corrected effect estimate and the number of putative missing studies. The corrected effect estimate is usually lower than the original estimate. The 95% confidence interval is re-computed, taking the observed and the inserted studies into account. We tested the difference between the original and the corrected effect size estimate provided by Trim and Fill for significance by investigating whether the original still fell between the confidence limits of the bias-corrected effect estimate. Importantly, the number of imputed studies should not be regarded as the definite number of missing studies. The procedure should rather be seen as a sensitivity analysis of how robust the original effect size estimate is against a possible publication bias (Duval, 2005).

All data sets were analyzed using the “Comprehensive Meta-Analysis 2.2” software (Borenstein, Hedges, Higgins & Rothstein, 2005). The various methods for the detection of publication bias can be applied after entering either the primary studies´ effect size statistics or the raw data. The methods can be applied to each effect size statistic (mean differences, correlations or dichotomous metrics).

Results

Description of meta-analyses investigated

Nineteen meta-analyses provided data sets that fulfilled all inclusion criteria and qualified for re-analysis in the present study. Our re-analysis included 31 data sets from these meta-analyses, which were homogeneous and contained more than 6 primary studies.

We drew one data set from each of the following meta-analyses: Barbato and D´Avanzo (2008), Beltman, Voshaar & Speckens (2010), Bortolotti et al. (2008), Cuijpers et al. (2006b, 2007, 2008), Dennis (2005), Ekers et al. (2008), Haby, Donnelly, Corry & Vos (2006) and Himelhoch et al. (2007). We were able to include two data sets each from Cuijpers et al. (2006a, 2009), Lynch, Laws & McKenna (2010), Neumeyer-Gromen et al. (2004), Newton-Howes, Tyrer & Johnson (2006), Pampallona, Bollini, Tibaldi, Kupelnick & Munizza (2004) and Reinecke, Ryan & DuBois (1998). Three data sets were provided by De Maat, Dekker, Schoevers & De Jonghe (2007), and four by De Maat, Dekker, Schoevers & De Jonghe (2006). A complete list of all 31 data sets included in the present study is given in Table 1.

We decided not to include the post-treatment and follow-up effects that were additionally reported in the 19 meta-analyses, because we found that these effect sizes did not fulfill our inclusion criteria.

Seven meta-analyses dealt with the efficacy of cognitive behavioral therapy (CBT). In three of them, the efficacy of CBT for depression was compared to control groups (Haby et al., 2006; Lynch et al., 2010; Reinecke et al., 1998). Two further meta-analyses dealt with the comparison of behavioural therapy (BT) to CBT (Cuijpers et al., 2007; Ekers et al., 2008). In one meta-analysis, the efficacy of CBT was compared for individual versus group therapy (Cuijpers et al., 2008). Finally, the efficacy of CBT in reducing comorbid depression in

individuals with somatic diseases compared to control groups was investigated by Beltman et al. (2010).

The efficacy of psychotherapy combined with pharmacotherapy was compared to pharmacotherapy alone by Pampallona et al. (2004), or psychotherapy alone by De Maat et al. (2007). In a meta-analysis by De Maat et al. (2006), psychotherapy was compared to pharmacotherapy. We will refer to combinations of different approaches in the same data set as psychotherapy (PT).

Six meta-analyses dealt with the efficacy of assorted psychotherapeutic approaches and for different samples. Cuijpers et al. (2006a) investigated the efficacy of PT for depressed adolescents, applied after a screening in school. The second meta-analysis investigated PT for older adults with latelife depression (Cuijpers et al., 2006b). The third meta-analysis dealt with the prevention and therapy of postnatal and postpartum depression (Dennis, 2005). Newton-Howes et al. (2006) compared PT for depressed participants with comorbid personality disorders to those without personality disorder. The efficacy of the "coping with depression" course as compared to control groups was assessed by Cuijpers et al. (2009). In a last meta-analysis, the efficacy of couple therapy was compared to individual therapy (Barbato & D´Avanzo, 2008).

PT can also be delivered to depressive patients in medical contexts, and to patients with somatic diseases and comorbid depression. Bortolotti et al. (2008) investigated the efficacy of PT delivered by medical practitioners in primary care, and Neumeyer-Gromen et al. (2004) investigated the efficacy of disease management programs (DMP), including psychoeducation, in reducing depression. The efficacy of PT in individuals with AIDS who suffer from comorbid depression was assessed by Himelhoch et al. (2007).

Unpublished studies were included in 5 of the 19 meta-analyses (Beltman et al., 2010; Cuijpers et al., 2006b; Cuijpers et al., 2008; Ekers et al., 2008; Neumeyer-Gromen et al., 2004). A statistical assessment of publication bias had been performed in 5 of the meta-analyses: a failsafe N was conducted in 2 (Cuijpers et al., 2008; Reinecke et al., 1998), Egger´s regression method was applied in 3 (Ekers et al., 2008; Himelhoch et al., 2007; Pampallona et al., 2004), and Beggs´s rank correlation test was performed in 1 meta-analysis (Himelhoch et al., 2007). All results of Begg´s and Egger´s tests were non-significant. In five meta-analyses funnel plots were used to detect asymmetry, albeit not always for the data sets that we included (Beltman et al., 2010; Ekers et al., 2008; Himelhoch et al., 2007; Neumeyer-Gromen et al., 2004; Newton-Howes et al., 2006). Some results of these visual examinations of the funnel plots suggested asymmetries which were discussed in the original analyses, but only Ekers et al. (2008) and Himelhoch et al. (2007) also used statistical tests, with non-significant results.

Results of the Trim and Fill calculations

Our recalculations of effect sizes using the raw data matched the original ones in most of the 31 data sets included with minor deviations in only a few cases (by .01 for 3 of the mean differences *d*, by .02 and by .04 for 1 of the mean differences *d,* respectively; by .01 for 3 relative risk (*RR*), by .02 for 2 *RR*, and by .03 for 1 *RR*; and lastly by .01 and by .03 for 1 of the odds ratios (*OR*), respectively). These small discrepancies were probably the result of different rounding. The re-calculated effect sizes were used for the statistical assessment of publication bias. The original and re-calculated effect sizes are shown in Table 1.

According to the Trim and Fill calculations, at least one missing study was identified in each of 12 data sets, reported by Ekers et al. (2008), Pampallona et al. (2004), De Maat et

al. (2006), Cuijpers et al. (2006a, 2006b, 2008, 2009), Newton-Howes et al. (2006), Himelhoch et al. (2007) and Bortolotti et al. (2008). We calculated corrected mean effect size estimates for all asymmetric data sets with the result that 11 of the 31 effect sizes were somewhat reduced. One of the corrected effect size estimates was actually enlarged. However, none of these changes in the effect sizes was significant, and in all but four cases all therapeutic approaches were assessed as significant both before and after correction. However, 2 effect size estimates that were non-significant before and after the correction both indicated an equal efficacy of 2 therapeutic approaches investigated, and 2 effect sizes changed from significance to non-significance, also indicating equal efficacy of the involved treatments after the correction for publication bias. Detailed information about all asymmetric data sets is presented below. Table 1 displays results of the Trim and Fill calculations.

Insert Table 1 about here

Table 1 shows the estimates of the efficacy of CBT for depression and reports 9 data sets (#1- #9), provided in 7 meta-analyses. Two of these data sets (#6, #8) showed evidence of publication bias, and additional studies were therefore imputed by the Trim and Fill procedure. In data set #6, 2 missing studies were imputed by the Trim and Fill procedure, but no significant reduction of the effect size resulted and the effect size differed significantly from zero both before and after correction. In data set #8, 3 studies were found to be missing, with no significant reduction in effect size. The pooled effect estimate was non-significant both before and after the reduction. No studies had to be imputed in the remaining data sets (#1- #5, #7, #9).

Missing studies were found in 2 of 9 data sets estimating the efficacy of PT alone compared to pharmacotherapy or a combined treatment of PT and pharmacotherapy to

pharmacotherapy. Four studies were found to be missing in data set #11, but firstly the reduction of the effect size by the Trim and Fill procedure was non-significant, and secondly both the original and the corrected effects size were non-significant. Three studies were found to be missing in data set #12, and the reduction was non-significant. The original effect size differed significantly from zero, but the corrected effect size indicated equal efficacy of PT and pharmacotherapy. No missing studies had to be imputed in the other data sets (#10, #13 - #18).

From 6 meta-analyses investigating the efficacy of various psychotherapeutic approaches for an assortment of subsamples with depressive symptoms, 9 data sets fulfilled the inclusion criteria of the present study. Of these, six were found to be asymmetric. The Trim and Fill procedure imputed 5 missing studies in data set #19, and 4 in #20, but the correction did not reduce the effect sizes significantly. Furthermore, both corrected effect sizes remained significant. In data set #21, 2 missing studies were imputed by the Trim and Fill procedure, but again there was no significant change in effect size. Both the original and the corrected effect size estimate differed significantly from zero. Data set #22 needed no correction.

OR larger than 1 indicated that treatment was more efficacious for depressed samples without comorbid personality disorder, than with personality disorder, in the two data sets #23 and #24. Effect size estimates in both data sets were reduced by the Trim and Fill procedure, indicating that 5 (data set #23) and 1 (#24) primary studies with non-significant effects were missing, but no significant differences between the original and the corrected effect size estimates occurred. However, in data set #23 the effect size turned non-significant after reduction, indicating that PT for patients with and without comorbidity is equally efficacious. Both the original and the reduced effect estimate in data set #24 differed significantly from

zero, and thus PT and pharmacotherapy for depressed patients without comorbidity are more efficacious. No missing studies were found in data set #25. In contrast, the pooled effect estimate of data set #26 was corrected by the Trim and Fill procedure. But even though eight missing studies had to be imputed, the resulting reduction of mean effect size was not significant. Effect size remained significant. Lastly, no missing studies had to be imputed in data set #27 (see Table 1).

Two data sets from the 3 meta-analyses investigating the efficacy of therapeutic interventions in medical contexts were affected by missing studies. The corrected effect size estimate of data set #28 was actually larger after adding one study in favor of PT, but there was no significant difference. This correction did not change the assessment of the efficacy of PT. No studies were missing in data sets #29 and #30, indicating that they are not affected by publication bias. In data set #31, 3 studies were found to be missing, but with no significant reduction of effect size, and both the original and corrected effect size differed significantly from zero.

Figure 1 shows the funnel plot for the data set with the largest number of studies estimated to be missing (N = 8; data set #26), investigating the reduction of depressive symptoms due to therapy with the "coping with depression" course (Cuijpers et al., 2009).

Insert Figure 1 about here

Results of the rank correlation and regression method

Using the rank correlation method (Begg & Mazumdar, 1994), one-sided testing indicated significant asymmetry in 9 data sets, and in 5 data sets using two-sided testing. The 9 asymmetric data sets found under one-sided testing (#7, #8, #9, #13, #19, #20, #23, #26 and #31) refer to the comparison of activity scheduling versus control conditions regarding the

reduction of depressive symptoms (#7), a comparison of the efficacy of BT versus CT/CBT in reducing depressive symptoms (#8), the comparison of CBT versus control groups for the reduction of depressive symptoms in participants with somatic diseases and comorbid depression (#9), the efficacy of PT compared to pharmacotherapy in reducing remission rates (#13), two data sets investigating the efficacy of psychotherapeutic treatment for adolescents after screening in schools (#19 & #20), a comparison of therapy for depressed samples with and without comorbid personality disorder (#23), therapy with the "coping with depression" course (#26), and the efficacy of group psychotherapy compared to control conditions in reducing depressive symptoms in samples with a HIV infection and comorbid depression (#31). The five asymmetric data sets significant under two-sided testing (#9, #13, #19, #20 and #26) were also significant under one-sided testing.

The linear regression method (Egger et al., 1997) identified a significant relationship between effect sizes and their variances in 9 data sets (one-sided testing): in data set #7, #8[1], #9, #19, #20, #23, #26 and #31, as was the case with Begg and Mazumdar´s method; and for data set #12, investigating the efficacy of PT compared to pharmacotherapy in reducing dropout rates. In 6 data sets, publication bias was found in two-sided testing (#9, #19, #20 and #26) as was the case in the rank correlation test, and in #12 and #23, as was the case for one-sided testing of the regression analysis (see Table 1).

Concordance between the three methods

[1] In the meta-analysis data set #8 was drawn from (Ekers et al., 2008), a non-significant intercept for Egger´s regression analysis was reported. Even though we used exactly the same data, we arrived at a different intercept by applying the software Comprehensive Meta-Analysis. We did not find an explanation for this discrepancy.

In the majority of the 31 assessed data sets (21 cases; data sets #1-6, #10, #11, #14-18, #21, #22, #24, #25, #27-30), none of the methods significantly indicated the presence of publication bias. Begg and Mazumdar's and Egger's tests were not significant in any of these 21 data sets, and the occasional imputation of missing studies according to the Trim and Fill procedure never resulted in a significant difference between the original and the corrected mean effect size estimates. Thus, there was high concordance between the three methods for the majority of cases. However, in 10 data sets either Begg and Mazumdar´s or Egger´s method indicated publication bias. When tested for significance one-sided, both methods led to the same conclusions for 8 data sets, for 4 sets when tested two-sided. Of the 10 data sets were either Begg and Mazumdar ´s or Egger´s methods indicated bias, the Trim and Fill procedure found missing studies in 7 cases, but none of the reductions was significant. However, in two of these data sets, the effect size changed from significance to non-significance after the correction. Thus, the efficacy of the respective interventions was changed by the publication bias, but the corrected effect sizes indicated an equal efficacy of two treatments in one of these data sets and equal efficacy of PT for depressed patients with and without comorbid PD in the other data set. In summary, we can therefore conclude that in most of the data sets where some evidence for publication bias was found, the efficacy of psychotherapy for depression and of preventive interventions was unchanged by the selective publication of positive results, and even in those data sets where it was changed, the efficacy of the interventions remained quite substantial.

Discussion

The present study aimed at investigating publication bias in meta-analyses of the efficacy of psychotherapeutic and preventive interventions for depression. An inappropriate application of the methods to detect publication bias was avoided by restricting the analysis to

homogeneous data sets including at least six primary studies (Egger et al., 1997; Ioannidis & Trikalinos, 2007). We found 31 data sets which fulfilled these inclusion criteria. Our results indicated only slight bias in most of these data sets. Begg and Mazumdar´s rank correlation method (1994) and Egger´s regression analysis (Egger et al., 1997) both indicated publication bias in 9 data sets (29.03%) for one-sided testing, and in 5 (16.13%) vs. 6 (19.35%) for two-sided testing. Both methods find essentially the same evidence for selective publication of positive results. In 12 data sets, including 7 of the data sets for which either of these methods indicated bias, the Trim and Fill procedure (Taylor & Tweedie, 2000a, 2000b) imputed missing studies, but the reduction of the pooled effect size estimates never reached significance. Moreover, the imputation of presumably missing studies did only twice result in a significant change of the effect estimates. The change of the respective effect sizes was not large itself, but as their significance changed, psychotherapy can no longer be regarded as more efficacious than pharmacotherapy after taking publication bias into account. The efficacy of both treatments needs to be regarded as equal. Furthermore, psychotherapy has to be considered as effective for depressed patients with comorbid PD as it is for those without PD. Therefore, in both cases we can conclude that the corrections due to publication bias did not yield inefficacious treatments. However, these changes in the efficacy of the respective interventions need to be considered when applying the treatments in clinical practice, and the corrected effect size estimates should be used as an orientation. In summary, it is a very important result that despite a minor tendency towards a selective publication of positive results, the efficacy of all reviewed interventions remains substantial even after correction using the Trim and Fill procedure. These results demonstrate that publication bias alone cannot explain their considerable efficacy. It can be concluded that despite some evidence for

publication bias, CBT and other psychotherapeutic interventions can still be considered efficacious and recommended for the treatment of depression.

Our results, taking heterogeneity into account, are considerably different from those of Cuijpers, Smit, et al. (2010), who found a large amount of publication bias in almost all of their data sets with a strong impact on mean effect size estimates. As we have pointed out, homogeneity is a necessary prerequisite of the application of statistical methods to detect publication bias, and tests for publication bias are likely to results in false positives in the presence of heterogeneity (Ioannidis & Trikalinos, 2007). The results of Cuijpers, Smit, et al. (2010) might be attributable to false positive hits, as the alternative explanation of heterogeneity cannot be ruled out for their results. The large differences between the original and the corrected effect sizes in the analysis of Cuijpers, for example the significant reduction of the overall mean effect size estimate from $g = .42$ to $g = .67$, with 51 imputed studies, is in strong contrast to the much smaller amounts of missing studies and effect reductions we have found. The difference between the results of the two analyses of publication bias is likely due to the emphasis on homogeneity in the present study. In contrast to Cuijpers, Smit, et al. (2010), we thus conclude that the efficacy of psychotherapy for depression is not strongly overestimated if the methods are applied to homogeneous data sets, as their assumptions require.

Importantly, our analysis did also differ from that of Cuijpers, Smit, et al. (2010), who did not re-calculate published meta-analyses, but conducted a new meta-analysis. Therefore, the sample of primary studies included in our re-analysis differs from that in Cuijpers, Smit, et al. (2010). The overlap varies from 0 to 12 studies between the 19 meta-analyses we included in the present study and Cuijpers, Smit, et al.´s analysis. The small amount of overlap is not surprising, since we aimed at including all meta-analyses for psychotherapeutic and

preventive interventions for depression. However, while the results of our analysis of publication bias might differ from that of Cuijpers, Smit, et al. (2010) due to the different pool of studies, it is unlikely that the degree of publication bias is so strongly discrepant between the two samples of included studies that this alone would explain since the differences.

Moreover, Cuijpers, Smit, et al. (2010) included only published studies in their analyses, whereas some unpublished studies were included in the data sets we assessed, albeit only few. Merely three of the included data sets comprehended unpublished studies. Data set #6 included 1 unpublished study (Cuijpers et al., 2008), data set #8 2 unpublished studies (Ekers et al., 2008), and #21 1 unpublished study (Cuijpers et al., 2006b). The inclusion of these four unpublished studies may have contributed only little to the relatively small tendency towards publication bias we found. Publication bias might be higher in samples with only published studies, but it is unlikely that this explains the difference in the results of the two analyses. In summary, however, even if the primary studies differ between the two analyses and the sample in which high publication bias was found included exclusively published studies (Cuijpers, Smit, et al., 2010), it is very unlikely that this can explain the large publication bias they found, in contrast to the present results.

Lastly, the significant results of Begg´s and Egger´s methods in our analysis were dispersed over all approaches. Also the occasional imputation of missing studies by the Trim and Fill procedure occurred in different approaches. A small tendency for selective reporting of positive outcomes seems to affect the different approaches quite equally, and no intervention is especially prone to the selective publication of positive results.

Two well-known methods developed for the detection of publication bias were not applied in the present study. First, the failsafe N (Becker, 2005) was not implemented because

it is of limited utility for a number of reasons[2]. Initially it was developed by Rosenthal (1979), the first to develop a statistical method to assess publication bias. He must be credited that he drew researcher's attention to the problem of publication bias. However, as the failsafe N method by Rosenthal and also the variants that have been developed by other statisticians all suffer from major drawbacks, more informative methods should be utilized instead (Becker, 2005). The second method is based on complex selection models (Hedges & Vevea, 2005) and describes the probability of publishing a study given its outcome. A weighted function of the p-values associated with each primary study, or a function of the primary studies´ effect size estimates and their associated standard errors, is used for the estimation of both the number of missing studies and their effect on the mean effect size estimate. As selection models are very difficult to conduct (Hedges & Vevea, 1996, 2005), they were also not applied in the present study.

To summarize the efficacy of the interventions under investigation, we can conclude that CBT is efficacious in reducing depressive symptoms in both adults and adolescents, in patients with somatic disease, and is also efficacious in reducing relapse. BT and CBT/CT are equally efficacious in reducing depressive symptoms, and activity scheduling is of higher efficacy than control conditions regarding the same outcome. Individual therapy is robustly more efficacious than group therapy in reducing depressive symptoms, but group psychotherapy is of higher efficacy in reducing depressive symptoms than control conditions.

---

[2] It focuses on statistical significance rather than clinical significance, the underlying assumption (Rosenthal, 1979) is a mean effect size of null in the missing studies, the sample size of the missing studies is not considered (studies with n=10 are set equal to studies with n=1000), nor are any study characteristics. No underlying statistical model exists, it is based on significance tests that combine p-values across studies (the current practice is to compute a p-value for the combined effect), and the distribution of the failsafe N statistic is unknown. Different formulas for the computation (Orwin, 1983; Fisher, 1932; Gleser & Olkin, 1936) lead to different estimates of the failsafe N, and none solve all of the problems mentioned (Becker, 2005).

Furthermore, couples and individual therapy are equally efficacious. The combined treatment of PT and antidepressants is more efficacious in in enhancing response rates and equally efficacious reducing dropout rates, compared to an antidepressant drug treatment alone. PT and combined treatment are equally efficacious in reducing dropout rates, and the combined treatment is of higher efficacy in enhancing remission rates compared to PT alone. PT alone is equally efficacious to pharmacotherapy in enhancing remission rates, and more efficacious in reducing dropout rates as well as relapse rates. PT for latelife depression, as well as for adolescents, can be considered efficacious in reducing depressive symptoms. However, PT is no more efficacious than control groups in preventing postnatal depression, and it is of equal efficacy for depressed patients with and without comorbid PD. PT and pharmacotherapy are robustly more efficacious in patients without comorbid PD. The "coping with depression" course is efficacious for the prevention and the therapy of depression. Lastly, DMP is more effective than primary care in reducing depressive symptoms, and PT in medical contexts is of higher efficacy than general practitioner care for the same outcome of interest. To conclude, a multitude of efficacious psychotherapeutic interventions does exist for depression, despite of publication bias (for all results see Table 1).

A limiting factor in our study is that only a few meta-analyses could be included in our assessment, as the necessary raw data needed for a post-hoc assessment of publication bias was not available for most of the meta-analyses. The reporting of the primary studies´ effect statistics and measures of their precision would be helpful for future investigations. Statistical guidelines such as the "Meta-analysis reporting standards" (MARS) included in the 6th edition of the APA Publication Manual (MARS; APA, 2008, 2009) and the "Preferred reporting items for systematic reviews and meta-analyses" (Prisma; Moher, Liberati, Tetzlaff

& Altmann, 2009) are available good standards that will help any researcher conducting a meta-analysis. Both recommend providing effect size statistics.

82.35% of the meta-analyses of psychotherapy and prevention for depression did not consider publication bias in their analyses, and 81.18% did not include unpublished studies. Therefore, it is necessary to emphasize the importance of assessing publication bias as a sensitivity analysis routinely, as is also recommended by the MARS and Prisma guidelines. This would solve the problem of missing data for follow-up analyses by other researchers. The MARS and Prisma guidelines should be mandatory in psychotherapeutic research. In addition, when methods for the detection of publication bias based on symmetry assumptions are applied, necessary statistical conditions for their employment must be met, in order to avoid erroneous interpretations of their results (Ioannidis & Trikalinos, 2007).

In summary, the positive conclusion that can be drawn from the present analysis is that publication bias does not seem to have invalidated the results of most of the meta-analyses of the efficacy of therapeutic approaches to treat depression. For the most part publication bias was either not present or did not change the conclusions. Thus, efficacious preventive and psychotherapeutic interventions for depression are available, and their efficacy does not seem to be based on a reporting bias. However, the correction for publication bias resulted in non-significance of effect sizes twice, and the modified evidence base for the respective interventions should be accounted for when applying them in clinical practice.

Acknowledgement

The authors would like to thank Helen-Rose Cleveland for language editing and Doris Heinrichs and Annett Schmitz for helpful technical assistance.

References

Meta-analyses are marked with an asterisk (*)

American Psychiatric Association (1980). *Diagnostic and Statistical Manual of Mental Disorders* (*3rd ed.*) (DSM-III). Washington, DC: American Psychiatric Association.

American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders (3rd ed. rev.).* Washington, DC: American Psychiatric Association.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders DSM-IV (4th ed.).* Washington, DC: American Psychiatric Association.

American Psychiatric Association (2000). *Diagnostical and statistical manual of mental disorders: DSM-IV-TR (4th ed., rev.).* Washington: DC: American Psychiatric Association.

American Psychological Association (2009). *Publication manual of the American Psychological Association (6th ed.)*. Washington: DC American Psychiatric Association.

* Andersson, G. & Cuijpers, P. (2009). Internet-based and other computerized psychological treatments for adult depression: A meta-analysis. *Cognitive Behaviour Therapy, 38*, 196-205.

APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008). Reportings standards for research in psychology. Why do we need them? What might they be? *American Psychologist*, *43,* 839-851.

* Barbato, A. & D'Avanzo, B. (2008). Efficacy of Couple Therapy as a Treatment for Depression: A Meta-Analysis. *Psychiatric Quarterly, 79*, 121-132.

Becker, B.J. (2005). Failsafe N or the file-drawer number. In: Rothstein, H.R., Sutton A.J., Borenstein M. (Eds.), *Publication bias in meta-analysis* (pp. 111-126). West Sussex: Wiley.

Begg, C.B. & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*, 1088-101.

* Bell, A.C. & D´Zurilla, T.J. (2009). Problem-solving therapy for depression: A meta-analysis. *Clinical Psychology Review, 29*, 348-353.

*Beltman, M.W., Voshaar, R.C.O. & Speckens, A. E. (2010). Cognitive behavioural therapy for depression in people with a somatic disease: meta-analysis of randomized controlled trials. *The British Journal of Psychiatry, 197*, 11-19.

Berlin, J.A. & Ghersi, D. (2005). Preventing publication bias: Registries and prospective meta-analysis. In: Rothstein, H.R., Sutton A.J., Borenstein M. (Eds.), *Publication bias in meta-analysis* (pp. 35-48). West Sussex: Wiley.

Birmaher, B., Ryan, N.D., Williamson, D.E., Brent, D.A., Kaufman, J., Dahl, R.E., Perel, J. & Nelson, B. (1996). Childhood and adolescent depression, I: a review of the past 10 years. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 1427-1439.

*Bledsoe, S.E. & Grote, N. (2006). Treating depression during pregnancy and the postpartum: A preliminary meta-analysis. *Research on Social Work Practice, 16*, 109-120.

Blumle, A, Antes, G., Schumacher, M., Just, H. & Von Elm, E. (2008). Clinical research projects at a German medical faculty: follow-up from ethical approval to publication and citation by others. *Journal of Medical Ethics, 34*, e20.

* Bohlmeijer, E., Smit, F. & Cuijpers, P. (2003). Effects of reminiscence and life review on late-life depression: A meta-analysis. *International Journal of Geriatric Psychiatry, 18*, 1088-1094.

Bosmans, J.E., Van Schaik, D.J., De Bruijne, M.C., Van Hout, H.P, Van Marwijk, H.W., Van Tulder, M.W. & Stalman, W.A. (2008). Are psychological treatments for depression in primary care cost-effective? *The Journal of Mental Health Policy and Economics, 11,* 3-15.

Borenstein M., Hedges L.V., Higgins J.P.T. & Rothstein H.R. (2005). *Comprehensive Meta-Analysis, Version 2.2.* Englewood, NJ: Biostat.

Borenstein M., Hedges L.V., Higgins J.P.T. & Rothstein H.R. (2009). *Introduction to meta-analysis*. West Sussex: Wiley & Sons.

* Bortolotti, B., Menchetti, M., Bellini, F., Montaguti, M.B. & Berardi, D. (2008). Psychological interventions for major depression in primary care: a meta-analytic review of randomized controlled trials. *General Hospital Psychiatry, 30*, 293-302.

Chalmers, I. (1990). Underreporting research is scientific misconduct. *The Journal of the American Medical Association, 263*, 1405-1408.

* Chan, E.K.-H. (2006). Efficacy of cognitive-behavioral, pharmacological, and combined treatments of depression: A meta-analysis. *Dissertation Abstracts International, Section B: The Sciences and Engineering, 67*, 2218.

* Cuijpers, P. (1997). Bibliotherapy in unipolar depression: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 28*, 139-147.

* Cuijpers, P. (1998). Psychological outreach programs for the depressed elderly: A meta-analysis of effects and dropout. *International Journal of Geriatric Psychiatry, 13*, 41-48.

* Cuijpers, P. (1998). A psychoeducational approach to the treatment of depression: a meta-analysis of Lewinsohns´s "Coping with Depression" course. *Behavior Therapy, 29*, 521-533.

* Cuijpers, P., Brannmark, J.G. & Van Straten, A. (2008). Psychological treatment of postpartum depression: A meta-analysis. *Journal of Clinical Psychology, 64*, 103-118.

* Cuijpers, P., Li, J., Hofmann, S.G. & Andersson, G. (2010). Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: A meta-analysis. *Clinical Psychology Review, 30*, 768-778.

* Cuijpers, P., Munoz, R.F., Clarke, G.N. & Lewinsohn, P.M. (2009). Psychoeducational treatment and prevention of depression: the “coping with depression” course thirty years later. *Clinical Psychology Review, 29*, 449-458.

* Cuijpers, P., Smit, F., Bohlmeijer, E., Hollon, S.D. & Andersson, G. (2010). Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. *The British Journal of Psychiatry*, *196*, 173-178.

* Cuijpers, P., Smit, F. & Van Straten, P. (2007). Psychological treatments of subthreshold depression: a meta-analytic review. *Acta Psychiatrica Scandinavica, 115*, 434-441.

* Cuijpers, P., Van Straten, A., Andersson, G. & Van Oppen, P. (2008). Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology, 76*, 909-922.

* Cuijpers, P., Van Straten, A., Hollon, S.D. & Andersson, G. (2010). The contribution of active medication to combined treatments of psychotherapy and pharmacotherapy for adult depression: A meta-analysis. *Acta Psychiatrica Scandinavica, 121*, 415-423.

* Cuijpers, P., Van Straten, A, Schuurmans, J., Van Oppen, P., Hollon, D. & Andersson, G. (2010). Psychptherapy for chronic major depression and dysthymia: A meta-analysis. *Clinical Psychology Review, 30*, 51-62.

* Cuijpers, P., Van Straten, A. & Smit, F. (2006b). Psychological treatment of late-life depression: A meta-analysis of randomized controlled trials. *International Journal of Geriatric Psychiatry, 21*, 1139-1149.

* Cuijpers, P., Van Straten, A., Smit, F., Mihalopoulos, & Beekman (2008). Preventing the onset of depressive disorders: a meta-analytic review of psychological interventions. *The American Journal of Psychiatry, 165*, 1272-1280.

* Cuijpers, P., Van Straten, A., Smits, N., & Smit, F. (2006a). Screening and early psychological intervention for depression in schools. Systematic review and meta-analysis. *European Child and Adolescent Psychiatry, 15*, 300-307.

* Cuijpers, P., Van Straten, A. & Warmerdam, L. (2007). Behavioral activation treatments of depression: A meta-analysis. *Clinical Psychology Review, 27*, 318-326.

* Cuijpers, P., Van Straten, A. & Warmerdam, L. (2007). Problem solving therapies for depression: A meta-analysis. *European Psychiatry, 22,* 9-15.

* Cuijpers, P., van Straten, A., Warmerdam, L. (2008). Are individual and group treatments equally effective in the treatment of depression in adults? A meta-analysis. *The European Journal of Psychiatry, 22*, 38-51.

* Cuijpers, P., Van Straten, A., Warmerdam L. & Smits, N. (2008). Characteristics of effective psychological treatments of depression: a metaregression analysis. *Psychotherapy Research*, 18, 225-236.

Degkwitz, R., Helmchen, H., Kockott, G. & Mombour, W., Hrsg. (1980). *Diagnosenschlüssel und Glossar psychiatrischer Krankheiten: dt. Ausgabe d. internat. Klassifikation d.*

Krankheiten d. WHO, ICD, 9. Revision [International Classification of Diseases, Ninth Revision]. Chapter V. Springer, Berlin, Heidelberg, New York.

* De Maat, S.M., Dekker, J., Schoevers, R.A. & De Jonghe, F. (2006). Relative efficacy of psychotherapy and pharmacotherapy in the treatment of depression: A meta-analysis. *Psychotherapy Research, 16*, 562-672.

* De Maat, S.M., Dekker, J., Schoevers, R.A. & De Jonghe, F. (2007). Relative efficacy of psychotherapy and combined therapy in the treatment of depression: A meta-analysis. *European Psychiatry, 22*, 1-8.

* Dennis, C.-L. (2005). Psychosocial and psychological interventions for prevention of postnatal depression: systematic review. *British Medical Journal, 331*, 1-8.

Demyttenaere, K., Bruffaerts, R., Posada-Villa, J. et al. (2004). Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *The Journal of the American Medical Association, 291*, 2581–90.

Dickersin, K. (2005). Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 11-34). West Sussex: Wiley & Sons Ltd.

* Dobson, K.S. (1989). A meta-analysis of the efficacy of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 57*, 414-419.

* Driessen E., Cuijpers P., De Maat , S.C.M., Abbass, A.A., De Jonghe, F., Dekker, J.J.M. (2010). The efficacy of short-term psychodynamic psychotherapy for depression: a meta-analysis. *Clinical Psychology Review, 30*, 25-36.

Duval, S. (2005). The Trim and Fill method. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 127-144). West Sussex: Wiley & Sons Ltd.

* Eifert, G.H. & Crail, L. (1989). The relationship between affect, behaviour, and cognition in behavioural and cognitive treatments of depression and phobic anxiety. *Behaviour change, 6*, 96-103.

* Ekers, D., Richards, D. & Gilbody, S. (2008). A meta-analysis of randomized trials of behavioural treatment of depression. *Psychological Medicine, 38*, 611-623.

Egger, M., Davey Smith, G., Schneider, M. & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629-34.

* Engels, G.I. & Vermey, M. (1997). Efficacy of Nonmedical Treatments of Depression in Elders: A Quantitative Analysis. *Journal of Clinical Geropsychology, 3*, 17-35.

Fisher, R.A. (1932). *Statistical methods for research workers* (4th Ed.). London: Oliver & Boyd.

* Friedman, M.A., Detweiler-Bedell, J. B., Leventhal, H.E., Horne, R., Keitner, G.I. & Miller, I.W. (2004). Combined psychotherapy and pharmacotherapy for the treatment of major depressive disorder. Clinical Psychology. *Science and Practice, 11*, 47-68.

* Gaffan, E.A., Tsaousis, I. & Kemp-Wheeler, S.M. (1995). Researcher allegiance and meta-analysis: The case of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 63*, 966-980.

* Gellatly, J., Bower, P., Hennessy, S., Richards, D., Gilbody, S. & Lovell, K. (2007).What makes self-help interventions effective in the management of depressive symptoms? Meta-analysis and meta-regression. *Psychological Medicine, 37*, 1217-1228.

* Gensichen, J., Beyer, M., Muth, C., Gerlach, F.-M., von Korff, M. & Ormel, J. (2006). Case management to improve major depression in primary health care: A systematic review. *Psychological Medicine, 36*, 7-14.

* Gerson, S., Berlin, T.R., Kaufman, A., Mintz, J. & Jarvik, L. (1999). Pharmacological and psychological treatments for depressed older patients: A meta-analysis and overview of recent findings. *Harvard Review of Psychiatry, 7*, 1-28.

Gilbody, S.M. & Song, F. (2000). Publication bias and the integrity of psychiatry research. *Psychological Medicine, 30*, 253- 258.

Glass, G.V., McGaw, B. & Smith, M.L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage Publications.

Gleser, L.J. & Olkin, I. (1996). Models for estimating the number of unpublished studies. *Statistics in Medicine*, *15*, 2493-2507.

* Gloaguen, V., Cottraux, J., Cucherat, M. & Blackburn, I.M. (1998). A meta-analysis of the effects of cognitive therapy in depressed patients. *Journal of Affective Disorders, 49*, 59-72.

* Gregory, R.J., Schwer Canning, S., Lee, T.W. & Wise, J.C. (2004). Cognitive bibliotherapy for depression: A meta-analysis. *Professional Psychology: Research and Practice, 35*, 275-280.

* Haby, M.M., Donnelly, M., Corry, J. & Vos, T. (2006). Cognitive behavioural therapy for depression, panic disorder and generalized anxiety disorder: A meta-regression of factors that may predict outcome. *Australian and New Zealand Journal of Psychiatry, 40*, 9-19.

* Haby, M.M., Tonge, B., Littlefield, L., Carter, R. & Vos, T. (2004). Cost-effectiveness of cognitive behavioural therapy and selective serotonin reuptake inhibitors for major

depression in children and adolescents. *Australian and New Zealand Journal of Psychiatry, 38,* 579-591.

Hedges, L.V. & Vevea, J. (1996). Estimating effect size under publication bias: Small sample properties and robustness of a random effects selection model. *Journal of Educational and Behavioral Statistics, 21*, 299-332.

Hedges, L.V. & Vevea, J. (2005). Selection method approaches. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 145-174). West Sussex: Wiley & Sons Ltd.

* Himelhoch, S., Medoff, D. & Oyeniyi, G. (2007). Efficacy of group psychotherapy to reduce depressive symptoms among HIV-infected individuals: A systematic review and meta-analysis. *Aids Patient Care and STDs, 21,* 732-739.

Hirschfeld, R. M. A., Keller, M. B., Panico, S., Arons, B. S., Barlow, D. H., Davidoff, F., et al. (1997). The National Depressive and Manic-Depressive Association consensus statement on the undertreatment of depression. *Journal of the American Medical Association, 277,* 333–340.

Hopewell, S., Clarke, M. & Mallett, S. (2005). Grey literature and systematic reviews. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), Publication bias in meta-analysis (pp. 49-72). West Sussex: Wiley & Sons Ltd.

* Imel, Z.E., Malterer, M.B., McKay, K.M. & Wampold, B.E. (2008). A meta-analysis of psychotherapy and medication in unipolar depression and dysthymia. *Journal of Affective Disorders, 110*, 197-206.

Ioannidis, J.P.A. & Trikalinos, T.A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *Canadian Medical Association Journal*, *176,* 1091-1096.

Ioannidis, J.P.A. (2005). Differentiating biases from genuine heterogeneity: Distinguishing artifactual from substantive effects. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 145-174). West Sussex: Wiley & Sons Ltd.

Jackson, D. (2006). The implications of publication bias for meta-analysis´ other parameter. *Statistics in Medicine, 25*, 2911-2921.

* Jane-Llopis, E., Hosman, C., Jenkins, R. & Anderson, P. (2003).Predictors of efficacy in depression prevention programs. *British Journal of Psychiatry, 183*, 384-397.

* Kang-Yi, C.D. & Gellis, Z.D. (2010). A systematic review of community-based health interventions on depression for older adults with heart disease. *Aging & Mental Health, 14*, 1-19.

* Klawansky, S. (1994). Meta-analysis on the treatment of depression in late life. In: Schneider, L.S., Reynolds, C.F., Lebowitz, B.D. & Friedhoff, A.J. (Eds.), *Diagnosis and treatment of depression in late life: Results of the NIH Consensus Development Conference* (pp. 333-352). Washington, DC, US: American Psychiatric Association.

* Koder, D.A., Brodaty, H. & Anstey, K.J. (1996). Cognitive therapy for depression in the elderly. *International Journal of Geriatric Psychiatry, 11*, 97-107.

* Kühner, C. (2003). Das Gruppenprogramm "Depression bewältigen" und seine Varianten – eine aktualisierte Meta-Analyse [The Coping with Depression Course and ist variants-an updated meta-analysis]. *Verhaltenstherapie, 13*, 254-262.

* Leichsenring, F. (2001). Comparative effects of short-term psychodynamic psychotherapy and cognitive-behavioral therapy in depression: A meta-analytic approach. *Clinical Psychology Review, 21,* 401-419.

Light, R.J. & Pillemer, D.B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

Lipsey, M.W. & Wilson, D.B. (2001). *Practical meta-analysis*. Thousand Oakes: Sage Publications Ltd.

* Lynch, D., Laws, K.R. & McKenna, J. (2010). Cognitive behavioural therapy for major psychiatric disorder: does it really work? A meta-analytical review of well-controlled trials. *Psychological Medicine, 40*, 9-24.

* Mazzuchelli, T., Kane, R. & Rees, C. (2009). Behavioral activation treatments for depression in adults: A meta-analysis and review. *Clinical Psychology: Science and Practice, 16*, 383-411.

* McCusker, J., Cole, M., Keller, E., Bellavance, F. & Berard, A. (1998). Effectiveness of treatments of depression in older ambulatory patients. Archives of Internal Medicine, 158, 705-712.

* McDermut, W., Miller, I.W. & Brown, R.A. (2001).The efficacy of group psychotherapy for depression: A meta-analysis and review of the empirical research. *Clinical Psychology: Science and Practice, 8*, 98-116.

McGauran, N., Wieseler, B., Kreis, J., Schüler, Y.B., Kölsch, H. & Kaiser, T. (2010). Reporting bias in medical research- a narrative review. *Trials, 11*, 1-15.

* Michael, K.D. (2000). The efficacy of treatments for childhood depression: An integrative review. *Dissertation Abstracts International, Section B: The Sciences and Engineering, 60,* 5784.

* Michael, K.D. & Crowley (2002). How effective are treatments for child and adolescent depression? A meta-analytic review. *Clinical Psychology Review, 22*, 247-269.

Moher, D., Liberati, A., Tetzlaff, J. & Altman, D.G. (2009). Preferred reporting items for systematic reviews and meta-analysis: The PRISMA statement. *Annals of Internal Medicine*, *151*, 264-270.

* Mohr, D.C. & Goodkin, D.E. (1999). Treatment of depression in multiple sclerosis: Review and meta-analysis. *Clinical Psychology: Science and Practice, 6*, 1-9.

Moscati, R., Jehle, D., Ellis, D., Fiorello, A. & Landi, M. (1994). Positive-outcome bias: Comparison of emergency medicine and general medicine literatures. *Academic Emergency Medicine, 1*, 267-271.

Moshagen, M. & Musch, J. (2008). Publication bias in studies on the efficacy of hypnosis as a therapeutic tool. *Contemporary Hypnosis*, *25*, 94-99.

* Nelson, M.K. (2002). Meta-analysis: Hypnotherapy/cognitive-behavioral therapy and its efficacy on depression compared to pharmacotherapy. *Dissertation Abstracts International, Section B: The Sciences and Engineering, 63,* 1040.

* Neumeyer-Gromen, A., Lampert, T., Stark, K. & Kallischnigg, G. (2004). Disease management programs for depression: A systematic review and meta-analysis of randomized controlled trials. *Medical Care, 42,* 1211-21.

* Newton-Howes, G., Tyrer, P. & Johnson, T. (2006). Personality disorder and the outcome of depression: Meta-analysis of published studies. *British Journal of Psychiatry, 188*, 13-20.

* Nietzel, M.T., Russell, R.L., Hemmings, K.A. & Gretter, M.L. (1987). Clinical significance of therapy for unipolar depression: A meta-analytic approach to social comparison. *Journal of Consulting and Clinical Psychology, 55*, 156-161.

* Oei, T.P.S. & Free, M.L. (1995). Do cognitive behaviour therapies validate cognitive models of mood disorders? A review of the empirical evidence. *International Journal of Psychology, 30*, 145-180.

Orwin, R.G. (1983). A fail-safe N for the effect size in meta-analysis. *Journal of Educational Statistics*, *8,* 157-159.

* Pampallona, S., Bollini, P., Tibaldi, G., Kupelnick, B. & Munizza, C. (2004). Combined pharmacotherapy and psychological treatment for depression. *Archives of General Psychiatry, 61*, 714-719.

Pearn, J. (1995). Publication: an ethical imperative. *British Medical Journal, 310*, 1313-1315.

* Pinquart, M., Duberstein, P.R. & Lyness, J.M. (2007). Effects of psychotherapy and other behavioral interventions on clinically depressed older adults: A meta-analysis. *Aging and Mental Health, 11*, 645-657.

* Pinquart, M., Duberstein, P.R. & Lyness, J.M. (2006). Treatments for later-life depressive conditions: a meta-analytic comparison of pharmacotherapy and psychotherapy. *The American Journal of Psychiatry, 163*, 1493-1501.

* Pinquart, M. & Sörensen, S. (2001). How effective are psychotherapeutic and other interventions with older adults? A meta-analysis. *Journal of Mental Health and Aging, 7*, 207-243.

* Posternak, M.A. & Miller, I. (2001).Untreated short-term course of major depression: a meta-analysis of outcomes from studies using wait-list control groups. *Journal of Affective Disorders, 66,* 139-146.

Psychotherapy: Randomized controlled and comparative trials (Electronic Data Base) (2011). Department of Clinical Psychology, VU University Amsterdam. Retrieved from http://www.evidencebasedpsychotherapies.org/index.php?id=23.

* Reinecke, M.A., Ryan, N.E. & DuBois, D.L. (1998). Cognitive-behavioral therapy of depression and depressive symptoms during adolescence: A review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry, 37*, 26-34.

* Robinson, L.A., Berman, J.S. & Neimeyer, R.A. (1990). Psychotherapy for the treatment of depression: a comprehensive review of controlled outcome research. *Psychological Bulletin, 108*, 30-49.

Rothstein, H.R., Sutton, A.J. & Borenstein, M. (2005). Publication bias in meta-analysis. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 1-8). West Sussex: Wiley & Sons Ltd.

Rosenthal, R. (1979). The „file drawer problem“ and tolerance for null results. *Psychological Bulletin*, *86*, 638-641.

* Schening, L.J. (2004). Refractory depression treatments options: A meta-analysis. *Dissertation Abstracts International, Section B: The Sciences and Engineering, 64,* 5233.

* Scogin, F. & McElreath, L. (1994). Efficacy of psychosocial treatments for geriatric depression: A quantitative review. *Journal of Consulting and Clinical Psychology, 62*, 69-74.

* Sin, N.L. & Lyubomirsky, S. (2009). Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: A practice-friendly meta-analysis. *Journal of Clinical Psychology, 65, 467-487.*

Singleton, N., Bumpstead, R., O'Brien, M., Lee, A. & Meltzer, H. (2001). Psychiatric morbidity among adults living in private households, 2000. *International Review of Psychiatry, 15,* 65-73.

Smith, M.L. (1980). Publication bias and meta-analysis. *Evaluation in Education: An International Review Series, 4,* 22-24.

* Steinbrück, S.M., Maxwell, S.E. & Howard, G.S. (1983). A meta-analysis of psychotherapy and drug therapy in the treatment of unipolar depression with adults. *Journal of Consulting and Clinical Psychology, 51*, 856-863.

Sterling, T.D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance- or vice versa. *Journal of the American Statistical Association*, *54,* 30-34.

Sterne, J.A.C., Becker, B.J. & Egger, M. (2005). The funnel plot. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 75-98). West Sussex: Wiley & Sons Ltd.

Sterne, J.A.C. & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (Eds.), *Publication bias in meta-analysis* (pp. 99-110). West Sussex: Wiley & Sons Ltd.

Sterne, J.A.C., Egger, M. & Moher, D. (2008). Addressing reporting biases. In: Higgins, J.P.T. & Green, S. (Eds.), *Cochrane handbook for systematic reviews of interventions*, pp. 297-334. Chichester: Wiley.

Sterne, J.A.C., Gavaghan, D. & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology, 53,* 1119-1129.

* Souter, M.A. & Miller, M.D. (2007). Do animal-assisted activities effectively treat depression? A meta-analysis. *Anthrozoos, 20*, 167-180.

* Spek, V., Cuijpers, P., Nyklicek, I., Riper, H., Keyzer, J. & Pop, V. (2007). Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: A meta-analysis. *Psychological Medicine, 37*, 319-328.

* Stuart, S. & Bowers, W.A. (1995). Cognitive therapy with inpatients: Review and meta-analysis. *Journal of Cognitive Psychotherapy, 9*, 85-92(8).

Sutton, A.J., Duval, S.J., Tweedie, R.L., Abrams, K.R. & Jones, D.R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *British Medical Journal*, *320,* 1574-1577.

* Taylor, T.L. & Montgomery, P. (2007). Can cognitive-behavioral therapy increase self-esteem among depressed adolescents? A systematic review. *Children and Youth Services Review, 29*, 823-839.

Taylor S. & Tweedie R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95,* 89-98.

Taylor S. & Tweedie R (2000b). Trim and Fill: A simple funnel-plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455-63.

Terrin, N., Schmid, C.H., Lau, J. & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine*, *22*, 2113–2126.

Vickers, A., Gloyal, N., Harland, R. & Rees, R. (1998). Do certain countries produce only positive results? A systematic review of controlled trials. *Controlled Clinical Trials, 19,* 159-166.

* Vittengl, J.R., Clark, L.A., Dunn, T.W. & Jarrett, R.B. (2007). Reducing relapse and recurrence in unipolar depression: A comparative meta-analysis of cognitive-

behavioral therapy´s effects. *Journal of Consulting and Clinical Psychology, 75*, 475-488.

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

* Wagner, J.J. (2005). A meta-analysis/ literature review comparing the effectiveness of SSRI antidepressants, cognitive behavioral therapy, and placebo for the treatment of depression. *Dissertation Abstracts International, Section B: The Sciences and Engineering, 66,* 2319.

* Wampold, B.E., Minami, T., Baskin, T.W. & Tierney, S.C. (2002). A meta-(re)analysis of the effects of cognitive therapy versus „other therapies“ for depression. *Journal of Affective Disorders, 68*, 159-165.

* Watanabe, N., Hunot, V., Omori, I.M., Churchill, R. & Furukawa, T.A. (2007). Psychotherapy for depression among among children and adolescents: a systematic review. *Acta Psychiatrica Scandinavica, 116*, 84-95.

* Weisz, J.R., McCarty, C.A. & Valeri, S.M. (2006). Effects of psychotherapy for depression in children and adolescents: A meta-analysis. *Psychological Bulletin, 132*, 132-149.

* Westen, D. & Morrison, K. (2001). A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology, 69,* 875-899.

*Wexler, B.E. & Cicchetti, D.V. (1992). The outpatient treatment of depression. Implications of outcome research for clinical practice. *Journal of Nervous and Mental Disease, 180*, 277–286.

World Health Organisation (1992). *The ICD-10 classification of mental and behavioral disorders: Clinical descriptions and diagnostic guidelines*. Geneva: Author.

Table 1: Results of the Trim and Fill procedure, Begg and Mazumdar's rank-correlation test, and Egger's regression analysis.

Data set number	Author	Intervention/ dependent measure/ time of measurement	Original effect size (95% CI)	Replicated effect size (95% CI)	Corrected effect size [+] (95% CI)	Number of studies included	Number of studies missing [++]	Begg (Tau) and Egger (β_0)
1	Haby et al. (2006)	CBT vs. control/symptom, functioning & quality of life combined/post	g = 0.54[a] (0.29; 0.79) (RE)	0.54[a] (0.29; 0.79)		11	0	Tau = .2 β_0 = -.4
2	Lynch et al. (2010)	CBT vs. control/depressive symptoms (HAMD)/post	g = -0.28[a] (-0.45; -0.12) (FE)	-0.28[a] (-0.44; -0.12)		9	0	Tau = .17 β_0 = 1.12
3	Lynch et al. (2010)	CBT vs. control/relapse/follow-up	OR = 0.53[a] (0.4; 0.72) (FE)	0.53[a] (0.4; 0.71)		9	0	Tau = -.93 β_0 = -.17
4	Reinecke et al. (1998)	CBT vs. control/depressive symptoms/adolescents/post	g = -1.02[a] (-1.23; -0.81)[b]	-1.02[a] (-1.22; -0.82) (FE)		6	0	Tau = -.19 β_0 = -.77
5	Reinecke et al. (1998)	CBT vs. control/depressive symptoms/adolescents/follow-up	g = -0.61[a] (-0.88; -0.35)[b]	-0.61[a] (-0.86; -0.35) (FE)		6	0	Tau = -.18 β_0 = -.04
6	Cuijpers et al. (2008)	Individual vs. group therapy/ depressive symptoms/post	d = 0.20[a] (0.05; 0.35) (RE)	0.20[a] (0.06; 0.35)	0.17[a] (0.01; 0.33)	15	2	Tau = .23 β_0 = .83
7	Cuijpers et al. (2007)	Activity scheduling vs. control group/depressive symptoms/ post	d = 0.87[a] (0.6; 1.15) (FE)	0.87[a] (0.6; 1.2)		10	0	Tau = .47* β_0 = 3.21*
8	Ekers et al. (2008)	BT vs. CT or CBT/depressive symptoms/post	d = 0.08 (-0.14; 0.3) (RE)	0.07 (-0.11; 0.24)	-0.02 (-0.19; 0.15)	12	3	Tau = .34* β_0 = 1.86*

Data set number	Author	Intervention/ dependent measure/ time of measurement	Original effect size (95% CI)	Replicated effect size (95% CI)	Corrected effect size [+] (95% CI)	Number of studies included	Number of studies missing [++]	Begg (Tau) and Egger (β_0)
9	Beltman et al. (2010)	CBT vs. control/somatic disease & depression/depressive symptoms/post	d = -0.16[a] (-0.27; -0.06) (FE)	-0.18[a] (-0.28; -0.08)		16	0	Tau = -.43*† β_0 = -2.12*†
10	Pampallona et al. (2004)	PT & antidepressant vs. antidepressant/responders/n.s.	OR = 1.86[a] (1.38; 2.52) (RE)	1.86[a] (1.38; 2.51)		16	0	Tau = 0.00 β_0 = .71
11	Pampallona et al. (2004)	PT & antidepressant vs. antidepressant/dropout/n.s.	OR = 0.86 (0.60; 1.24) (RE)	0.86 (0.60; 1.24)	0.72 (0.49; 1.06)	16	4	Tau = .17 β_0= .78
12	De Maat et al. (2006)	PT vs. pharmacotherapy/drop-out/post	RR = 1.29[a] (1.07; 1.57) (FE)	1.27[a] (1.05; 1.54)	1.19 (0.99; 1.42)	10	3	Tau = .38 β_0 = 1.07*†
13	De Maat et al. (2006)	PT vs. pharmacotherapy/ remission/post	RR = 0.91 (0.79; 1.06) (FE)	0.93 (0.8; 1.08)		10	0	Tau = -.6*† β_0 = -1.51
14	De Maat et al. (2006)	PT vs. pharmacotherapy/ remission/post	OR = 0.87 (0.68; 1.10) (FE)	0.9 (0.72; 1.18)		10	0	Tau = -.22 β_0 = -.23
15	De Maat et al. (2006)	PT vs. pharmacotherapy/relapse/ follow-up	RR = 0.46[a] (0.33; 0.65) (FE)	0.49[a] (0.35; 0.69)		6	0	Tau = -.33 β_0 = -.68
16	De Maat et al. (2007)	PT vs. PT & pharmacotherapy/ drop-out/post	RR = 1.03 (0.82; 1.3) (FE)	1.04 (0.83 ; 1.31)		7	0	Tau = -.05 β_0 = .27
17	De Maat et al. (2007)	PT vs. PT & pharmacotherapy/ remission/post	RR = 1.32[a] (1.12; 1.56) (FE)	1.33[a] (1.13; 1.56)		7	0	Tau = -.05 β_0 = -1.48
18	De Maat et al. (2007)	PT vs. PT & pharmacotherapy/ remission/post	OR = 1.59[a] (1.22; 2.09) (FE)	1.6[a] (1.22; 2.1)		7	0	Tau = -.33 β_0 = -1.07

Data set number	Author	Intervention/ dependent measure/ time of measurement	Original effect size (95% CI)	Replicated effect size (95% CI)	Corrected effect size [+] (95% CI)	Number of studies included	Number of studies missing [++]	Begg (Tau) and Egger (β_0)
19	Cuijpers et al. (2006a)	Screening & PT/depressive symptoms/post	d = 0.58[a] (0.37; 0.78) (FE)	0.59[a] (0.39; 0.79)	0.41[a] (0.23; 0.59)	8	5	Tau = .73*† β_0 = 2.27*†
20	Cuijpers et al. (2006a)	Screening & PT/depressive symptoms/one study less/post	d = 0.72[a] (0.45; 0.99) (FE)	0.76[a] (0.51; 1.02)	0.56[a] (0.33; 0.79)	7	4	Tau = .67*† β_0 = 3.54*†
21	Cuijpers et al. (2006b)	PT vs. control/latelife depression/depressive symptoms/post	d = 0.72[a] (0.57 ; 0.87) (FE)	0.72[a] (0.57; 0.87)	0.69[a] (0.54; 0.84)	14	2	Tau = .25 β_0 = .74
22	Dennis (2005)	PT vs. control/prevention of postnatal depression/depressive symptoms/post	g = -0.06 (-0.37; 0.26) (FE)	-0.06 (-0.37; 0.26)		7	0	Tau = -.14 β_0 = -.34
23	Newton-Howes et al. (2006)	PT/comorbid PD vs. no PD/ RCTs/depressive symptoms/ post	OR = 1.6[a] (1.25; 2.06) (RE)	1.61[a] (1.25; 2.06)	1.29 (0.98; 1.69)	14	5	Tau = .36* β_0 = 1.67*†
24	Newton-Howes et al. (2006)	PT & pharmacotherapy/ comorbid PD vs. no PD/ depressive symptoms/post	OR = 1.74[a] (1.25; 2.42) (RE)	1.74[a] (1.25; 2.41)	1.71[a] (1.23; 2.37)	10	1	Tau = .11 β_0 = .72
25	Cuijpers et al. (2009)	Coping with depression, prevention/depressive symptoms/post	RR = 0.62[a] (0.43; 0.91) (RE)	0.62[a] (0.43; 0.91)		6	0	Tau = -.2 β_0 = -1.21
26	Cuijpers et al. (2009)	Coping with depression, therapy/depressive symptoms/post	d = 0.28[a] (0.18; 0.38) (RE)	0.28[a] (0.19; 0.38)	0.19[a] (0.08; 0.3)	18	8	Tau = .29*† β_0 = 1.79*†

Data set number	Author	Intervention/ dependent measure/ time of measurement	Original effect size (95% CI)	Replicated effect size (95% CI)	Corrected effect size [+] (95% CI)	Number of studies included	Number of studies missing [++]	Begg (Tau) and Egger (β_0)
27	Barbato & D´Avanzo (2008)	Couple vs. individual therapy/ drop-out/post	RR = 1.16 (0.67; 2.01) (RE)	1.16 (0.67; 2.02)		6	0	Tau = .07 β_0 = -1.44
28	Bortolotti et al. (2008)	PT vs. GP/depressive symptoms/ post	d = -0.42[a] (-0.59; -0.26) (FE & RE)	-0.43[a] (-0.59; -0.26)	-0.46[a] (-0.62; -0.3)	6	1	Tau = .2 β_0 = .99
29	Neumeyer-Gromen et al. (2004)	DMP vs. usual primary care/ depressive symptoms/follow-up	RR = 0.75[a] (0.7; 0.81) (FE)	0.75[a] (0.7; 0.8)		10	0	Tau = -.09 β_0 = -.3
30	Neumeyer-Gromen et al. (2004)	DMP vs. usual primary care/ depressive symptoms/no minor depression/follow-up	RR = 0.74[a] (0.69; 0.8) (FE)	0.74[a] (0.7; 0.8)		10	0	Tau = -.2 β_0 = -1.07
31	Himelhoch et al. (2007)	Group PT vs. control/depression & HIV/depressive symptoms/ post	d = 0.38[a] (0.23; 0.53) (RE)	d = 0.38[a] (0.23; 0.53)	0.33[a] (0.17; 0.48)	8	3	Tau = .44* β_0 = 1.47*

Note. CBT = Cognitive behavioral therapy; CI = confidence interval; d = Cohen´s d; DMP = disease management program; FE = fixed effects model; g = Hedges´ g; GP = general practitioner care; OR = odds ratio; PT = psychotherapy; PD = personality disorder; r = correlation; RE = random effects model; RR = relative risk; [+] = for publication bias; [++] =according to Trim and Fill

[a] significant effect size with a 95% confidence interval not including zero (for RR and OR: confidence interval not including 1)

[b] the integration model was not specified in the paper

* $p < .05$, one-sided

† $p < .05$, two-sided

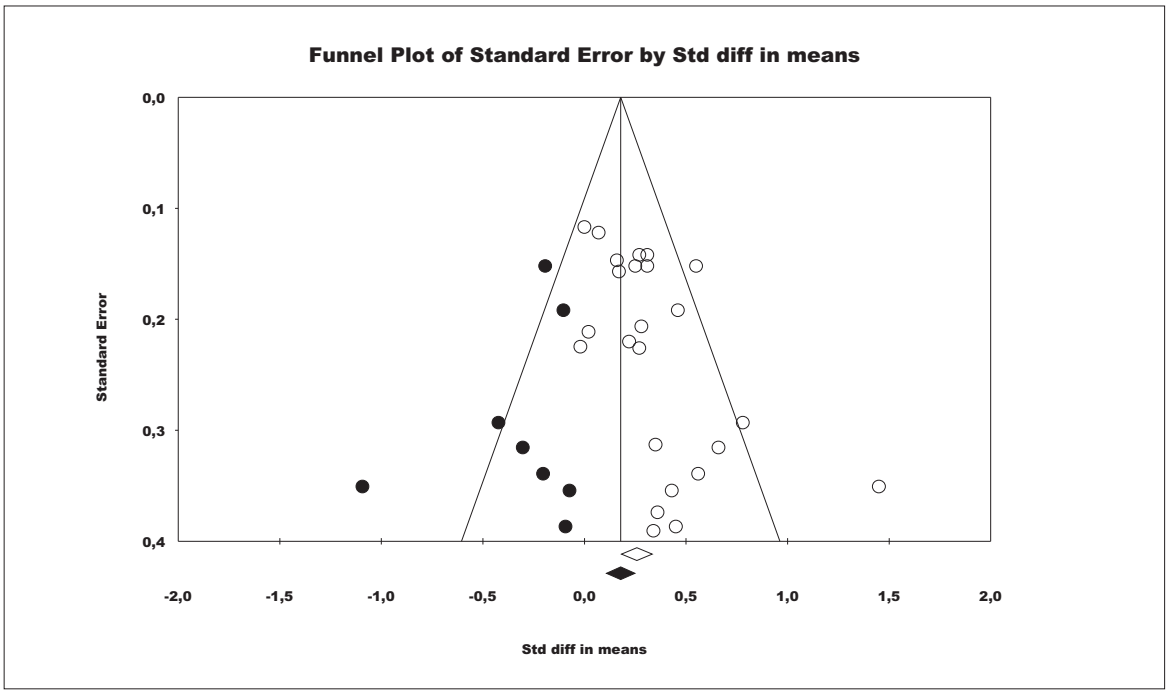

Figure 1.

An illustration of the Trim and Fill method showing the funnel plot for the reduction of depressive symptoms due to therapy with the "coping with depression" course (data set #26; Cuijpers et al., 2009). Dark circles: the 8 studies identified as missing.

Erklärung

Hiermit erkläre ich, daß ich die hier vorgelegte Dissertation eigenständig und ohne unerlaubte Hilfe angefertigt habe.

Die Dissertation wurde in der vorgelegten oder in ähnlicher Form noch keiner anderen Institution eingereicht. Ebenso versichere ich, daß bisher keine erfolglosen Promotionsversuche stattgefunden haben.

Düsseldorf, den 18.10.2011

(Helen Niemeyer)