



universität
wien

Diplomarbeit

Titel der Arbeit

Wo sind die Unterschiede? Das Dodo Bird Verdict und
die relative Effektivität von Bona Fide-Psychotherapien
für die Posttraumatische Belastungsstörung.

Eine Metaanalyse.

Verfasserin

Bettina Gregor

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im März 2012

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Mag. Dr. Ulrich Tran

Danksagung

Eine Diplomarbeit zu schreiben nimmt ein gutes Stück Lebenszeit in Anspruch. Ich möchte mich daher bei jenen Menschen bedanken, die dazu beigetragen haben, dass es eine lohnende Zeit war und dass sie zu einem guten Abschluss gekommen ist.

Allen voran gilt mein Dank meinem Betreuer, Mag. Dr. Ulrich Tran, der in mir das Interesse für Metaanalysen und das Dodo Bird Verdict geweckt hat und durch seinen persönlichen Einsatz und sein präzises und scharfsinniges Denken dafür gesorgt hat, dass dieses Interesse noch weiterhin besteht. Dass ich heute mit Stolz das Gefühl haben kann, mit meiner Arbeit einen kleinen Beitrag zu einer aktuellen wissenschaftlichen Diskussion geleistet zu haben, ist für mich persönlich von hohem Wert. Ich bedanke mich sehr für Ihre Unterstützung!

Da eine Metaanalyse vorzubereiten und zu rechnen ein einsames Unterfangen ist, gebührt mein Dank meiner Studienkollegin und Freundin Mag. Claudia Mies, deren emotionale Unterstützung mich gut durch die Hochs und Tiefs dieser Zeit gebracht hat. Sie ist mir mit Humor, aufmunternden Worten und guten Tipps zur Seite gestanden, wodurch ich nie das Gefühl hatte, mit meiner Arbeit ganz alleine zu sein.

Mein besonderer Dank gilt meinen Freundinnen, allen voran Mag. Karin Lechle und Mag. Ursula Schöff, die an mich geglaubt haben und deren vielfältigem Einsatz es zu verdanken ist, dass ich mein Studium wieder aufgenommen habe und es trotz so mancher Hindernisse auch zu Ende führe.

Ich bedanke mich auch bei meiner Familie und meinem Freundeskreis, die mir ihre Unterstützung nach ihren Möglichkeiten zukommen ließen.

Eine Person aber ist dafür verantwortlich, dass ich heute diese Danksagung schreiben kann. Ohne sie hätte ich vermutlich Vergleichende Sprachwissenschaft studiert und wieder abgebrochen, hätte in mehrfachem Sinne am Hungertuch genagt und viele Male das Studium wegen Unvereinbarkeit mit Beruf und Familie wieder aufgegeben. Mein Lebensgefährte, Mag. Franz Hellerschmid, hat in meiner Studienzeit seine Bedürfnisse zurückgestellt, um mir diese Chance im Leben zu ermöglichen. Mein Dank dafür lässt sich in Worten kaum ausdrücken. Dennoch: Ich danke dir!

Inhaltsverzeichnis

1	Einleitung	11
2	Theoretische Grundlagen	12
2.1	<i>Das Dodo Bird Verdict</i>	12
2.2	<i>Die Metaanalysen von Wampold</i>	15
2.3	<i>Die Kritik</i>	18
2.3.1	Sind alle psychotherapeutischen Behandlungsverfahren gleich effektiv?	18
2.3.2	Ist das Vorgehen in den relevanten Metaanalysen dazu geeignet, Unterschiede in der Wirksamkeit von Therapieverfahren sichtbar zu machen?	21
2.3.3	Welchen Anteil an der Varianz haben allgemeine und spezifische Faktoren?	24
2.3.4	Was bedeutet die äquivalente Wirksamkeit von Psychotherapieverfahren?	25
2.4	<i>Die Posttraumatische Belastungsstörung</i>	26
2.4.1	Das Störungsbild	27
2.4.2	Die Metaanalyse von Benish et al. (2008) und der Stand der Forschung	30
2.4.3	Kritik an der Bona Fide-Definition	32
3	Zielsetzung und Fragestellung	34
4	Studienauswahl und Suchstrategie	35
4.1	<i>Inklusions- und Exklusionskriterien</i>	35
4.1.1	Die Bona Fide-Definition	35
4.2	<i>Suchstrategie</i>	37
5	Kodierung	39
5.1	<i>PatientInnenanzahl</i>	40
5.2	<i>Variablen</i>	41
5.2.1	Primäre Maße	41
5.2.2	Sekundäre Maße	43
5.2.3	Follow-ups	45
5.2.4	Klinische Signifikanz	46
5.2.5	Dropouts	47

5.3	<i>Kategorisierung der Behandlungsverfahren</i>	47
5.3.1	EMDR	48
5.3.2	Expositionsverfahren und Prolonged Exposure	48
5.3.3	CBT	49
5.3.4	Gegenwartszentrierte Therapie (PCT, Present-Centered Therapy)	50
5.3.5	Traumafokussierte versus nicht-traumafokussierte Verfahren	50
5.4	<i>Moderatoren</i>	51
5.4.1	Studienqualität	52
5.5	<i>Fehlende Werte</i>	54
6	Statistische Methodik	55
6.1	<i>Effektgrößenberechnung</i>	55
6.1.1	Standardisierter Mittelwertsunterschied: Cohens <i>d</i> , Hedges <i>g</i>	55
6.1.2	Relatives Risiko	57
6.2	<i>Metaanalyse und statistisches Modell</i>	58
6.3	<i>Heterogenität</i>	60
6.4	<i>Metaanalytische Verfahren</i>	62
6.4.1	Random Signs-Verfahren	62
6.4.2	Upper Bound-Verfahren	63
6.4.3	Kategorienvergleiche	64
6.5	<i>Sensitivitätsanalysen</i>	64
6.5.1	Abhängigkeiten	64
6.5.2	Publication Bias	64
6.5.3	Einflussreiche Fälle	66
6.6	<i>Moderatoranalysen</i>	66
6.7	<i>Statistisches Programm</i>	68
7	Deskriptive Ergebnisse	69
7.1	<i>Neue Studien</i>	69
7.2	<i>Beschreibung aller inkludierten PTSD-Studien</i>	72
7.2.1	Studienstichprobe	72
7.2.2	PatientInnenstichprobe	74
7.2.3	Inkludierte Behandlungsverfahren	75
7.3	<i>Effektgrößen</i>	78
8	Ergebnisse der Metaanalysen	82
8.1	<i>Das Random Signs-Verfahren</i>	82
8.2	<i>Das Upper Bound-Verfahren</i>	85

8.3	<i>Kategorienvergleiche</i>	89
8.3.1	Primäre und sekundäre Maße	89
8.3.2	Detailanalysen zu den primären und sekundären Maßen	93
8.3.3	Sensitivitätsanalysen zu den primären und sekundären Maßen	95
8.3.4	Klinische Signifikanz	97
8.3.5	Dropouts	100
8.3.6	Moderatorenanalysen	102
9	Zusammenfassung und Diskussion	104
9.1	<i>Zusammenfassung der Hauptergebnisse</i>	104
9.2	<i>Sind Bona Fide-Psychotherapien gleich wirksam?</i>	106
9.2.1	Die gegenwartszentrierten Behandlungsverfahren	107
9.2.2	Traumafokussierte versus nicht-traumafokussierte Behandlungsverfahren	109
9.2.3	Die Bona Fide-Definition	110
9.2.4	Das Dodo Bird Verdict	111
9.3	<i>Kritik und Grenzen</i>	112
9.4	<i>Zukünftige Forschung</i>	114
10	Literaturverzeichnis	116
11	Tabellenverzeichnis	134
12	Abbildungsverzeichnis	135
13	ANHANG	137

Hinweis

Ich versichere, dass die vorliegende Diplomarbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt wurde. Des Weiteren bestätige ich, dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt wurde. Alle Ausführungen der Arbeit, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.



„Everybody has won, and *all* must have prizes.“

(Carroll, 1920, S. 34)

Abbildung 1. Illustration von John Tenniel (1820 – 1914) aus *Alice's adventures in wonderland* von Lewis Carroll. Der Schiedsspruch des Dodo Vogels. Copyright abgelaufen.

„Bei diesem Stand der Dinge nimmt es nicht wunder,
dass alle psychotherapeutischen Schulen sich halten können.
Trotz langer und harter Polemiken ist es bisher noch keiner Schule gelungen,
eine mit ihr rivalisierende aus dem Feld zu schlagen. Die naheliegende Folgerung ist,
dass wohl alle zu etwas gut sein müssen, dass aber noch keine zu Ergebnissen
gekommen ist, die denen einer anderen klar überlegen wären.“

Jerome D. Frank (1997, S. 427)

1 Einleitung

Die vorliegende Diplomarbeit befasst sich mit einem Themenbereich der Psychotherapieforschung. Sie versucht die Frage zu beantworten, ob verschiedene Psychotherapien gleich wirksam sind oder ob es Verfahren gibt, die effektiver sind als andere. Das Verhältnis der Wirksamkeit von Therapieverfahren zueinander wird als relative Effektivität bezeichnet.

Die psychotherapeutischen Verfahren, die Gegenstand dieser Untersuchung sind, werden zur Behandlung der Posttraumatischen Belastungsstörung eingesetzt. Diese psychische Störung kann nach dem Erleben von traumatischen Ereignissen wie Gewalterfahrungen oder Kriegserlebnissen auftreten. Der Begriff Psychotherapie, mit dem diese Verfahren bezeichnet werden, erfüllt nicht notwendigerweise die österreichischen gesetzlichen Anforderungen. Es handelt sich um etablierte psychologische Interventionen, deren Ziel es ist, die Symptome der traumatisierten PatientInnen zu lindern. Die Bezeichnungen dieser Interventionen wurden übersetzt, wenn dies eindeutig möglich war, teilweise wurden aber zur besseren Identifizierung die englischen Begriffe beibehalten.

Um die Tatsache zu würdigen, dass PatientInnen, TherapeutInnen oder StudienautorInnen sowohl Frauen als auch Männer sein können, wurde in diesen oder ähnlichen Begriffen das Binnen-I verwendet. Wenn aber die geschlechtsneutrale Schreibweise Formulierungen verkompliziert hätte, wurde darauf verzichtet und der männliche Begriff benutzt. Dennoch sind damit beide Geschlechter angesprochen.

Das methodische Verfahren, das in dieser Diplomarbeit verwendet wurde, ist das der Metaanalyse. Eine Metaanalyse bietet die Möglichkeit, die Ergebnisse vieler Einzelstudien statistisch zusammenzufassen und damit den Stand der Forschung quantitativ darzustellen. Zu Beginn dieser Arbeit erfolgt eine Darstellung der theoretischen Grundlagen der Thematik, woraus die Forschungsfragen abgeleitet werden. Anschließend werden die Methodik, nach der diese Metaanalyse durchgeführt wurde, ausführlich dargelegt und die Ergebnisse aller Analysen berichtet. Eine Zusammenfassung der wichtigsten Ergebnisse wird zum Schluss diskutiert und kritisch beleuchtet.

2 Theoretische Grundlagen

2.1 Das Dodo Bird Verdict

„At last the Dodo said, ‘*Everybody* has won, and *all* must have prizes’.“

Seit 1936 Saul Rosenzweig seinem Artikel „Some implicit common factors in diverse methods of psychotherapy“ dieses Zitat aus *Alice im Wunderland* vorangestellt hat, bezieht man sich in der Psychotherapieforschung mit dem Begriff *Dodo Bird Verdict* darauf, dass alle Psychotherapien gleich wirksam sind (Budd & Hughes, 2009; Hunsley & Di Giulio, 2002; Luborsky et al., 2002). Rosenzweig gab in dieser historisch bedeutungsvollen Arbeit eine Erklärung für diese Annahme, die ihn zum Begründer der *Common Factors* Theorie machte (Duncan, 2002). Er unterschied Elemente, von denen angenommen wird, dass sie in einer bestimmten Therapie wirken, von jenen Faktoren, die tatsächlich wirksam sind und die seiner Meinung nach alle psychologischen Therapien gemeinsam haben. Seine Arbeit hatte einen wesentlichen Einfluss auf spätere Vertreter dieser Theorie. Frank (1997) verwendete eine Metapher, um zwei mögliche Grundhaltungen zu veranschaulichen:

Zwei offenbar höchst verschiedene Psychotherapien, etwa Psychoanalyse und systematische Desensibilisierung, könnten sich zueinander entweder verhalten wie Penicillin zu Digitalis – zwei pharmakologisch völlig verschiedene Wirkstoffe, die sich für die Behandlung völlig verschiedener Leiden eignen. Oder aber das aktive Element in beiden könnte das gleiche sein, ähnlich wie wenn zwei Präparate mit verschiedenen Namen auf den Markt kommen, die beide Aspirin enthalten. Ich glaube, die zweite Möglichkeit kommt der Wahrheit näher. (S. 428)

Frank (1997) postulierte vier zentrale allgemeine Elemente von Psychotherapie: eine vertrauensvolle Beziehung zur helfenden Person, ein heilsames Setting, eine Behandlungstheorie, die eine plausible Erklärung für die Symptome und deren Heilung bietet, und ein Ritual beziehungsweise eine Vorgehensweise, an der sich PatientIn und TherapeutIn aktiv beteiligen und an dessen Wirksamkeit sie glauben (Duncan, 2002; Frank, 1997).

Obwohl einzelne Studien immer wieder signifikante Unterschiede in der Wirksamkeit von Behandlungsverfahren berichten und diese daher als Argument gegen das Dodo Bird Verdict herangezogen werden (DeRubeis, Brotman, & Gibbons, 2005; Westmacott & Hunsley, 2007), können nur metaanalytische Verfahren den Stand der Forschung quantifizieren und damit einen objektiven Beitrag zur Klärung dieser Hypothese leisten. Im Gegensatz zu einer narrativen Zusammenfassung und Bewertung von allen vorhandenen relevanten Einzelstudien bietet eine Metaanalyse eine statistische Synthese der Resultate dieser Studien. Analog zu den inkludierten Primärstudien werden ähnliche Fragestellungen untersucht und ähnliche statistische Werte wie Mittelwert und Standardabweichung benutzt, um zu diesen Fragestellungen eine Aussage zu machen (Borenstein, Hedges, Higgins, & Rothstein, 2009).

Mehr als 30 Jahre Forschungsarbeit sind seit der ersten Metaanalyse von Smith und Glass (1977) vergangen, die die Wirksamkeit verschiedener psychotherapeutischer Verfahren untersuchte. Diese Metaanalyse brachte den eindeutigen Beleg, dass Psychotherapie wirkt, fand aber nur geringe Unterschiede zwischen den einzelnen Verfahren. Obwohl die Ergebnisse der metaanalytischen Forschung seither tendenziell die Äquivalenzannahme stützen (Luborsky et al., 2002), ist diese Sichtweise sehr umstritten und wird in der psychotherapeutischen Fachwelt nur von wenigen akzeptiert (Budd & Hughes, 2009; Chambless, 2002; Hunsley & Di Giulio, 2002; Rounsaville & Carroll, 2002; Westmacott & Hunsley, 2007). Einer der Gründe dafür ist, dass in den Metaanalysen, deren Ergebnisse für das Dodo Bird Verdict sprechen (Luborsky et al., 2002; Smith & Glass, 1977; Wampold et al., 1997), sowohl unterschiedliche Behandlungsverfahren als auch unterschiedliche Störungsbilder integriert waren (Beutler, 2002). Dieses verallgemeinernde Vorgehen verdeckte möglicherweise Unterschiede zwischen den Verfahren ebenso wie das Aggregieren von unterschiedlichen Ergebnismaßen zu einer einzigen Effektgröße (Siev, Huppert, & Chambless, 2009). Darin waren störungsspezifische Maße genauso enthalten wie Maße, die nicht die primäre, sondern mögliche komorbide Störungen betrafen. Deshalb sind die empirischen Befunde, die für oder gegen das Dodo Bird Verdict sprechen, eng mit methodischen Fragen von Metaanalysen verknüpft.

Auch wenn letztlich die Evidenzlage noch ungeklärt ist (Tolin, 2010), sind vor allem die Auswirkungen auf die Praxis bedeutungsvoll. Als eine mögliche Antwort,

warum empirisch validierte Therapieformen, die vor allem aus der kognitiven Verhaltenstherapie (CBT, Cognitive-Behavioral Therapy) stammen, nicht ausreichend verbreitet sind, finden sich genau jene Ansichten unter PsychotherapeutInnen wieder: zum einen, dass alle Psychotherapien gleich effektiv sind und zum anderen, dass die allgemeinen Wirkfaktoren für den größten Anteil am Ergebnis von Therapiestudien verantwortlich sind. Um diesen Annahmen, die der Implementierung empirisch validierter Behandlungsverfahren in der Praxis entgegenstehen, wirkungsvoll entgegenzutreten, lud der Präsident der *Association for Behavioral and Cognitive Therapies* seine Mitglieder dazu ein, sich am wissenschaftlichen Diskurs zu beteiligen und betonte: „Either we rebut these conclusions, conduct new research to show they are wrong, or we accept them and change our message“ (DiGuiseppe, 2007, S. 119).

Gerade im Bereich der CBT besteht seit gut 50 Jahren das Forschungsinteresse darin, spezifische Interventionen für spezifische Aufgabenstellungen zu entwickeln und diese laufend zu verbessern. Damit wird eine Gegenposition zum Dodo Bird Verdict eingenommen, die lautet, dass es Therapieverfahren gibt, die für bestimmte Störungen aufgrund ihrer spezifischen Bestandteile und Techniken wirkungsvoller sind als andere (Budd & Hughes, 2009). Um die Metapher von Frank (1997) weiterzuführen: Digitalis bewährt sich bei Herzinsuffizienz besser als Penicillin. Deshalb bemühten sich die Forschergruppen der CBT intensiv darum zu zeigen, dass ihre Interventionen wirksamer sind als andere Behandlungsverfahren. Das führte bereits vor 25 Jahren zu einem Überhang an Psychotherapiestudien aus dem verhaltenstheoretischen Bereich (Shapiro & Shapiro, 1983). Es ist allerdings schwierig, fundierte Schlussfolgerungen über den Vergleich von Verfahren zu ziehen, wenn andere therapeutische Interventionen nicht oder kaum auf ihre Wirksamkeit hin überprüft werden (DiGuiseppe, 2007). Die Tatsache, dass es für bestimmte Störungen wesentlich mehr Wirksamkeitsstudien für CBT gibt, könnte auch mit den Therapieverfahren zusammenhängen. Unter Umständen wurden Interventionen anderer Therapierichtungen für ein spezifisches Störungsbild als nicht hilfreich erachtet, und daher wurde auf Wirksamkeitsüberprüfungen verzichtet (Chambless, 2002).

Nicht nur, dass psychotherapeutische Verfahren anderer Therapieschulen hinsichtlich ihrer Effektivität seltener untersucht wurden, auch die Motivation bei der Durchführung von Wirksamkeitsstudien scheint sich zu unterscheiden und eher

darin zu liegen, aufzuzeigen, dass ihre Verfahren genauso effektiv sind wie verhaltenstheoretische Interventionen (Budd & Hughes, 2009). In der Präsentation der Ergebnisse einer randomisierten kontrollierten Studie zur Effektivität von gestaltorientierter Traumatherapie auf der *12th European Conference on Traumatic Stress* (Butollo & Karl, 2011), sprach der Vortragende von einem Genozid der humanistischen Therapieverfahren in der Wirksamkeitsforschung (Butollo, 3.6.2011). Diese Aussage lässt auf die Übermacht verhaltenstheoretischer Forschung schließen, auch wenn der Bemerkung eine ironische Komponente anhaftete. Aufgrund dieser unterschiedlichen Positionen stammen die KritikerInnen des Dodo Bird Verdicts typischerweise aus dem Bereich der CBT. Wie kürzlich publizierten Beiträgen zu entnehmen ist, ist die Diskussion unter BefürworterInnen und KritikerInnen (wieder) aktuell (Carroll & Rounsaville, 2010; DiGuiseppe, 2007; Ehlers et al., 2010; Siev et al., 2009; Wampold et al., 2010), und es sollen nun in Folge die für die vorliegende Arbeit wichtigsten Argumente zusammengefasst werden.

2.2 Die Metaanalysen von Wampold

Ein Versuch, das Dodo Bird Verdict empirisch zu belegen und dabei drei typische methodische Probleme von Metaanalysen bei der Prüfung der relativen Effektivität zu umgehen, war die Metaanalyse von Wampold et al. (1997). „In what has perhaps become the *de facto* last word on the matter, ...“ (Tolin, 2010, S. 711), verwendeten sie nur Studien mit einem direkten Vergleich von mindestens zwei Therapieformen, um nach Shadish und Sweeney (1991) eine mögliche Konfundierung der Unterschiede in den Effektgrößen mit Unterschieden in den Variablen wie zum Beispiel Dauer der Behandlung oder Diagnosekriterien zu vermeiden. Weiters verzichteten sie auf die Kategorisierung von Behandlungsverfahren, um statistische Probleme aufgrund zahlreicher Paarvergleiche und Definitionsprobleme bei der Kategorisierung zu umgehen. Schließlich benutzten sie eine Definition von *Bona Fide-Therapien*, um Behandlungsverfahren auszuschließen, die nur als Kontrollbedingung und nicht als therapeutisches Verfahren (*intended-to-be-therapeutic*) gedacht waren. Ihre Hypothese, dass alle Bona Fide-Psychotherapien gleich effektiv sind, wurde mit einem eigenen Verfahren, einem *Omnibus-Test* zur Feststellung der Verteilungsform, geprüft und mit einem Standardverfahren zur Feststellung der Obergrenze des wahren Effekts ergänzt. Ergebnis war, dass die Effektgrößen in Übereinstimmung mit

der Nullhypothese homogen um Null verteilt waren und die Obergrenze des wahren Effekts bei $d = 0.19$ lag. Die Schlussfolgerung im Sinne des Dodo Bird Verdicts wurde schon im Titel des Artikels angeführt: „... Empirically, ‚All Must Have Prizes‘.“

Diese Metaanalyse war auch ein wichtiger empirischer Baustein von Wampolds *The Great Psychotherapy Debate* (2001), wo zwei der Psychotherapie zugrundeliegende Metamodelle einander gegenübergestellt wurden: Das medizinische Modell beruht darauf, dass vergleichbar mit der Medizin eine Störung korrekt diagnostiziert wird, es für diese Störung eine psychologische Erklärung und davon abgeleitet einen Mechanismus der Veränderung gibt, sodass spezifische, darauf abgestimmte Bestandteile eines Behandlungsverfahrens angewendet werden können, die dann für die Störung heilsam sind (Wampold, 2001, S. 13f.).

Das kontextabhängige Modell beruht auf den allgemeinen Wirkfaktoren und wurzelt in dem Werk *Persuasion and Healing* (Frank, 1997). Sechs Elemente werden als wesentliche allgemeine Faktoren genannt, die der Therapeut entlang des Prozesses für den Klienten nutzbar machen sollte: die therapeutische Beziehung, die Hoffnung auf eine Verbesserung, die Möglichkeit neuer Lernerfahrungen, das Hervorrufen von Emotionen, die Steigerung der Selbstwirksamkeitserwartung und die Möglichkeiten zum praktischen Üben und Trainieren (Wampold, 2001, S. 25). Gerade der Wert der spezifischen Techniken wurde von Wampold betont:

Specific ingredients are necessary to any bona fide psychotherapy whether conceptualized as a medical model treatment or a contextual model treatment. In the contextual model, specific ingredients are necessary to construct a coherent treatment that therapists have faith in and that provides a convincing rationale to clients. (S. 25)

Im medizinischen Modell liegt der Wert der spezifischen Techniken in ihrem psychologischen Effekt und im kontextabhängigen Modell in der Bedeutung, die PatientIn und TherapeutIn diesen Techniken verleihen. Für Wampold selbst war dies eine zentrale Frage seiner Arbeit: „If the two treatments are equally effective, is it because of their respective specific ingredients or because both are instances of contextual model treatments?“ (Wampold, 2001, S. 27).

Nach Sichtung der empirischen Datenlage kam er zu dem Schluss, dass das gängige medizinische Modell verworfen werden müsste. Ein wesentlicher Bestandteil der Argumentation war das Dodo Bird Verdict und damit das Fehlen der Unterschiede in der Wirksamkeit unterschiedlicher Behandlungsverfahren.

Allerdings wurde die Studie von 1997 unter anderem hinsichtlich ihrer Heterogenität und Generalisierbarkeit kritisiert und hinterfragt (Crits-Christoph, 1997; Howard, Krause, Saunders, & Kopta, 1997). Das Hauptargument war, dass eine heterogene Studienauswahl über Zielgruppen, Störungsbilder und Therapieformen hinweg eine Übergeneralisierung darstellt und mit der spezifischen Methodik des Omnibus-Tests Unterschiede zwischen den Behandlungsverfahren bei spezifischen Störungsbildern überdeckt werden könnten. Selbst wenn es keinen Unterschied zwischen allen Behandlungsverfahren für alle Störungsbilder gäbe, so bedeute das nicht, dass bei einem spezifischen Störungsbild nicht ein bestimmtes Behandlungsverfahren effektiver sein könnte als andere aktive Verfahren; so lautete das seitdem vielfach zitierte Argument (Beutler, 2002; Chambless, 2002; Crits-Christoph, 1997; Hunsley & Di Giulio, 2002; Siev et al., 2009). Exakt die Hälfte der inkludierten Studien betrafen Angstproblematiken und etwa 69% der Studien berichteten über Vergleiche von Behandlungsverfahren aus dem Bereich der CBT (Crits-Christoph, 1997). Insofern erschien es 1997 tatsächlich als nicht legitim und darüberhinaus verfrüht, weitreichende Schlussfolgerungen im Sinne des Dodo Bird Verdicts für alle Psychotherapien und alle Störungsbilder zu treffen, was übrigens auch Wampold et al. (1997) selbst als Einschränkung ihrer Metaanalyse anführten.

In weiterer Folge erschienen Metaanalysen auf der Basis des Designs von 1997, die sich mit zwei Störungsbildern, der Posttraumatischen Belastungsstörung (PTSD, Posttraumatic Stress Disorder; Benish, Imel, & Wampold, 2008) und alkoholinduzierten Störungen (Imel, Wampold, Miller, & Fleming, 2008), und einer spezifischen Zielgruppe, nämlich Kinder und Jugendlichen (Miller, Wampold, & Varhely, 2008) befassten. In all diesen Metaanalysen wurde die Nullhypothese bestätigt, dass die Effektgrößen aus den direkten Vergleichen der Bona Fide-Therapien nicht signifikant von Null abweichen, und dass Abweichungen mit *Researcher Allegiance* (der theoretischen Ausrichtung der ForscherInnen, die eine Verzerrung zugunsten einer Therapieform bewirken kann) erklärt werden können (Imel et al., 2008).

Ein wichtiger Grund für mögliche Unterschiede zu Gunsten einzelner Therapieformen in anderen Metaanalysen lag nach Wampold in Vergleichstherapien, die nur als Kontrollbedingung konzipiert waren. Wampold, Minami, Baskin und Tierney (2002) führten eine Reanalyse einer Metaanalyse durch, in der sich die Kognitive Therapie einer Mischkategorie von Therapien überlegen zeigte (Gloaguen, Cottraux, Cucherat, & Blackburn, 1998). Nachdem die Therapien dieser Kategorie in bona fide und nicht-bona fide entsprechend der Definition geteilt wurden, zeigte sich die Kognitive Therapie gleich wirksam wie die Bona Fide-Therapien, war aber weiterhin den als nicht-bona fide qualifizierten Therapien überlegen. Die Studienauswahl nach der Bona Fide-Definition wurde in der Kontroverse um die Ergebnisse der PTSD-Metaanalyse von Benish et al. (2008) thematisiert (Ehlers et al., 2010; Wampold et al., 2010) und wird unter 2.4.3 und 9.2.3 diskutiert.

2.3 Die Kritik

Die Hauptargumente der KritikerInnen lassen sich vier grundsätzlichen Fragestellungen zuordnen. Sind alle psychotherapeutischen Behandlungsverfahren tatsächlich gleich effektiv? Ist das Vorgehen in den relevanten Metaanalysen dazu geeignet, Unterschiede in der Wirksamkeit von Therapieverfahren sichtbar zu machen? Welchen Anteil an der Varianz haben allgemeine und spezifische Faktoren? Was bedeutet die äquivalente Wirksamkeit von psychotherapeutischen Verfahren?

2.3.1 Sind alle psychotherapeutischen Behandlungsverfahren gleich effektiv?

Die KritikerInnen des Dodo Bird Verdicts nennen in Bezug auf diese Frage naturgemäß die Gegenbeispiele. Sowohl in Primärstudien als auch in Metaanalysen fanden sich Unterschiede zwischen Therapieverfahren, vor allem bei bestimmten Störungsbildern. Gerade für die Behandlung von Angststörungen wurden zahlreiche Gegenbeispiele genannt (Chambless, 2002; DeRubeis et al., 2005). Dass aber eine einzelne Studie einen signifikanten Unterschied zwischen zwei Behandlungsverfahren auch aufgrund des Typ I Fehlers finden könnte, ist ein Argument, das die Forschergruppen um Wampold immer wieder betonten (Benish et al., 2008; Wampold et al., 1997). Daher fassten Siev und Chambless (2007) jeweils fünf Primärstudien zum Vergleich von Kognitiver Therapie und Entspannungstherapie für zwei Angststörungen zusammen und führten eine Metaanalyse durch. Sie konnten zwar keinen Unter-

schied in der Behandlung der Generalisierten Angststörung nachweisen, aber sehr wohl einen signifikanten Unterschied für die Panikstörung ohne Agoraphobie, was sie als Beweis für die Spezifität eines Behandlungsverfahrens für ein Störungsbild werteten. Auch für andere Störungsbilder gibt es einzelne Metaanalysen, die die größere Effektivität eines Behandlungsverfahrens belegen. Im Bereich der Depression ist nach einer Metaanalyse die Interpersonale Psychotherapie der CBT überlegen (De Mello, De Jesus Mari, Bacaltchuk, Verdelli, & Neugebauer, 2005), wobei allerdings nur drei Primärstudien in diese Analyse eingingen.

Somit müssen auch methodische Aspekte bei der Bewertung der relativen Effektivität berücksichtigt werden. War beispielsweise in den Primärstudien der Vergleich zwischen den Therapieverfahren fair? Wie bereits erwähnt, führt unter Umständen Researcher Allegiance zu einer ungleichen Ausgangslage für zwei Behandlungsverfahren innerhalb einer Studie. Ein deutlicher Vorteil für die kognitive Therapie zur Behandlung von Depressionen (Dobson, 1989) wurde nach Berücksichtigung von Researcher Allegiance wieder relativiert, wenn auch nicht zur Gänze (Gaffan, Tsaousis, & Kemp-Wheeler, 1995).

Ein weiteres Problem ist die Uneinheitlichkeit der Ergebnisse von verschiedenen Metaanalysen zu ein und derselben Therapieform. Beispielsweise kamen die Metaanalysen über die Psychodynamische Therapie von Depressionen zu unterschiedlichen Ergebnissen (Driessen et al., 2010; Leichsenring, 2001; Leichsenring & Rabung, 2008; Svartberg & Stiles, 1991). Auch innerhalb einer Metaanalyse können sich widersprüchliche Ergebnisse finden. Hier ist aus methodischer Sicht die Studie von Churchill et al. (2001) ein interessantes Beispiel, da Gruppenunterschiede bei dichotomen, nicht aber bei kontinuierlichen Ergebnisvariablen berichtet wurden.

Um eine Aussage zur relativen Effektivität von Behandlungsverfahren für ein Störungsbild vornehmen zu können, erscheint es daher notwendig, die Metaanalysen aus diesem Feld zu sichten und zu bewerten, wie dies Luborsky et al. (2002) für das Dodo Bird Verdict getan haben oder alle verfügbaren Wirksamkeitsstudien nach einer einheitlichen und fundierten metaanalytischen Methodik auszuwerten. Der Frage nach der relativen Effektivität von Psychotherapieverfahren wird in Metaanalysen typischerweise auf zweifache Weise begegnet. Häufig werden unterschiedlichste Primärstudien integriert, in denen aktive Behandlungsverfahren mit Kontroll-

gruppen verglichen werden und fallweise auch mit anderen aktiven Behandlungsverfahren. Die Effektgrößen, die aus diesen Vergleichen entstehen, werden dann zueinander in Beziehung gesetzt, und es wird daraus die Schlussfolgerung gezogen, welchen Therapieverfahren der Vorzug zu geben ist. Aber bereits 1991 wurde darauf hingewiesen, dass diese Resultate von vielen Faktoren beeinflusst sein können, die sich von Studie zu Studie unterscheiden (Shadish & Sweeney, 1991). Aus diesem Grund wurde nur der direkte Vergleich zweier aktiver Behandlungsverfahren empfohlen, um den Einfluss von Studiencharakteristika zu minimieren und ein reliables Ergebnis zur relativen Effektivität von Psychotherapieverfahren zu erhalten.

Abgesehen von den Forschergruppen um Wampold wählten auch Cuijpers, van Straten, Andersson und van Oppen (2008) diesen Ansatz für ihre Metaanalyse über Depressionsstudien. Sie untersuchten 53 Studien mit Direktvergleichen und konnten kaum Unterschiede zwischen den Therapien feststellen: „This study suggests that there are no large differences in efficacy between the major psychotherapies for mild to moderate depression“ (S. 909). Dieses Ergebnis wurde auch durch eine systematische Überprüfung von Metaanalysen und randomisierten, kontrollierten Studien zur Depressionsbehandlung gestützt, in der explizit methodisch höchste Qualität und Aktualität gefordert war. Dabei wurden nur Metaanalysen und Studien, die seit dem Jahr 2000 erschienen waren, in die Bewertung inkludiert. Zusammenfassend konnten keine Unterschiede zwischen den psychologischen Verfahren festgestellt werden (Berg & Høie, 2010), wenn man auch berücksichtigen muss, dass depressive PatientInnen auf unterschiedliche Therapien vergleichbarer ansprechen als dies bei anderen Störungsbildern der Fall ist (Beutler, 2002).

Eine aktuelle Metaanalyse leistet einen weiteren wichtigen Beitrag zu der Frage nach der relativen Effektivität von CBT im Vergleich zu anderen psychotherapeutischen Verfahren (Tolin, 2010). Darin wurden viele der kontroversen Aspekte der laufenden Diskussionen aufgegriffen: Es wurden nur Direktvergleiche von CBT mit anderen Bona Fide-Psychotherapien analysiert, Ergebnismaße wurden differenziert erfasst, die Studienqualität wurde nach zwei Verfahren beurteilt und ebenso als Moderator einbezogen wie eine Bewertung von Researcher Allegiance. Insgesamt 26 Studien wurden mit dem Ergebnis analysiert, dass CBT wirksamer ist als psychodynamische Verfahren, aber nicht wirksamer als interpersonale oder supportive

Verfahren. Die Überlegenheit der CBT galt allerdings nur für PatientInnen mit Angststörungen oder depressiver Symptomatik.

Auch wenn die Metaanalysen von Siev und Chambless (2007) und Tolin (2010) aufgrund ihrer methodischen Klarheit ein wichtiges Argument für die Spezifitätshypothese darstellen, so schlossen die Vertreter des Dodo Bird Verdicts diese Möglichkeit gar nicht aus. Auf Ebene der Primärstudien sahen Wampold et al. (1997) einzelne Studien, die signifikante Unterschiede zwischen Behandlungsverfahren berichteten, als seltene Abweichungen in der Verteilung der Effektgrößen um Null. Außerdem wurde einzelnen Methoden durchaus zugestanden, bei bestimmten Störungsbildern oder PatientInnengruppen größere Therapieeffekte zu bewirken. Diese werden als Ausnahmen betrachtet (Frank, 1997). Den BefürworterInnen des Dodo Bird Verdicts geht es um den Großteil der empirischen Befunde. Mit den Worten von Duncan (2002): „Besides the occasional significant finding for a particular therapy, the critical mass of data revealed no differences in effectiveness between the various treatments for psychological distress“ (S. 35). Den KritikerInnen hingegen geht es um die Gefahr, dass wichtige Unterschiede übersehen beziehungsweise gar nicht erforscht werden. Für seinen Fachbereich Depression formulierte dies Cuijpers (1998) so:

... but we have to be careful, because the 'Dodo Bird Verdict' can be criticised in many ways. First, and perhaps foremost, is that the Dodo gave his verdict much too early. Because of the verdict it seems as if we know everything we ought to know about the effects of psychotherapy for depression. That is absolutely not the case. (S. 357)

2.3.2 Ist das Vorgehen in den relevanten Metaanalysen dazu geeignet, Unterschiede in der Wirksamkeit von Therapieverfahren sichtbar zu machen?

Wie bereits erwähnt, zielte ein Teil der Kritik auf die Übergeneralisierung und Undifferenziertheit innerhalb der Metaanalysen, die als Beleg für das Dodo Bird Verdict gelten (Chambless, 2002).

It is now widely agreed that it is a mistake to ask whether any school of therapy is superior to any other school of therapy, on average, or for all disorders (cf. Shapiro & Shapiro, 1983). To ask such a question is akin to asking whether insulin or an antibiotic is better, without knowing the condition for

which these treatments are to be given. . . . Alternatively, researchers should begin with a problem and ask how treatments compare in their effectiveness for that problem. (DeRubeis et al., 2005, S. 175)

Dieser Kritikpunkt führte bei den Forschergruppen um Wampold dazu, dass Metaanalysen für spezifische Störungsbilder und eine spezifische Zielgruppe durchgeführt wurden (siehe 2.2), er spiegelt sich aber auch im Umgang mit den Ergebnissen der Primärstudien wider. Wenn alle Ergebnismaße gemittelt werden, um nur eine einzige Effektgröße zu haben, die in die Analyse eingeht, kann dieses Vorgehen bedeutende Unterschiede unsichtbar machen. Sowohl für die Forschung als auch für die Praxis ist es jedoch wichtig, zu erfahren, ob Kernsymptome einer Störung, deren diagnostischer Status und/oder die komorbide Symptomatik durch ein Verfahren verbessert wurden (Siev et al., 2009). In diesem Sinne wurde für eine Differenzierung in primäre und sekundäre Maße plädiert (Crits-Christoph, 1997). Wie sehr eine Übergeneralisierung Unterschiede verdecken kann, zeigte eine Analyse von Effektgrößen am Beispiel von unveröffentlichten Daten aus der Metaanalyse von Siev und Chambless (2007). Während die Effektgröße für den Unterschied von CBT und Entspannungsverfahren für beide Störungsbilder $d = 0.19$ war und damit mit der Obergrenze von Wampold et al. (1997) konform ging, änderte sich die Effektgröße substantiell, wenn sie nach Störungsbild differenziert wurde: Für die Panikstörung ergab sich ein signifikanter Unterschied zwischen den Therapieverfahren von $d = 0.31$, während der Unterschied für die generalisierte Angststörung mit $d = 0.08$ nicht signifikant blieb (Siev et al., 2009). Ebenso unterschieden sich die Effektgrößen für die primären und sekundären Maße für die Panikstörung deutlich, wenn sie getrennt erhoben wurden (primäre Maße $d = 0.49$, $p < .002$, sekundäre Maße $d = 0.02$, $p = .89$), während für die generalisierte Angststörung keine Unterschiede vorhanden waren.

Einem eher verallgemeinerndem Vorgehen stellen die KritikerInnen des Dodo Bird Verdicts daher eine differenzierende Herangehensweise gegenüber (Siev et al., 2009; Tolin, 2010):

What is the appropriate question to be asked of outcome research? In all its complexity, the question towards which all outcome research should ultimately be directed is the following: *What* treatment, by *whom*, is most

effective for *this* individual with *that* specific problem, and under *which* set of circumstances? (Paul, 1967, S. 111)

Im Bereich der alkoholinduzierten Störungen findet sich aktuell ein Zugang zu neuen Möglichkeiten in der Psychotherapieforschung, der durch eine groß angelegte Studie, die für die Therapie von Alkoholmissbrauch wegweisend war, ausgelöst wurde (Project MATCH, 1997). In diesem Projekt wurden drei unterschiedliche therapeutische Verfahren in ihrer Wirksamkeit verglichen und eine Reihe von differentiellen Hypothesen geprüft. Alle drei Gruppen zeigten größtenteils sehr ähnliche positive Resultate und die meisten differentiellen Hypothesen wurden nicht gestützt. So folgerte Willenbring in seinem Übersichtsartikel „The past and future of research on treatment of alcohol dependence“ (2010), dass sich einige Leitideen der Forschung auf diesem Gebiet als falsch oder unzureichend erwiesen haben, unter anderem die Annahme, dass sich technische Unterschiede von Therapien in unterschiedlichen Ergebnissen, zumindest für bestimmte PatientInnengruppen, niederschlagen würden.

In einer erst kürzlich publizierten Zusammenfassung des aktuellen Wissensstandes über Suchtverhalten wurde ebenfalls konstatiert, dass unterschiedliche Psychotherapien scheinbar ganz ähnliche Ergebnisse erzielen (Sellman, 2010). Auf der Suche nach einem alternativen Forschungsansatz schlugen Longabaugh und Morgenstern (1999) ein individualisiertes therapeutisches Vorgehen anhand von Entscheidungsbäumen vor, die an Stelle des therapeutischen Verfahrens Gegenstand der Forschung sein könnten. Insgesamt scheint es, als ob die Therapieforschung gerade im Bereich Alkoholmissbrauch aufgrund der schwachen Befunde für die Spezifität einzelner Verfahren bereit ist, neue Wege zu gehen: „For now we believe that it is important for our field to recognize fully that the specific effects hypothesis remains unproven and to consider the implications of this on future research directions“ (Morgenstern & McKay, 2007, S. 1385). Dennoch gibt es auch im Feld der alkoholinduzierten Störungen kritische Stimmen, die beispielsweise mit dem Argument effektiver computergestützter Therapien, die ohne therapeutische Beziehung funktionieren und einen eindeutig durch das Verfahren vermittelten Kompetenzzuwachs erzielen, darauf hinweisen, dass wichtige Unterschiede zwischen Verfahren sichtbar werden, wenn man systematisch danach sucht, und es auf keinen Fall egal ist, welche therapeutischen Verfahren eingesetzt werden (Carroll & Rounsaville, 2010).

2.3.3 Welchen Anteil an der Varianz haben allgemeine und spezifische Faktoren?

Der dritte Bereich, an dem sich die Kritik entzündet, ist der Anteil einzelner Faktoren an der Varianzaufklärung des Ergebnisses von Wirksamkeitsstudien, allen voran TherapeutInneneffekte (beziehungsweise die therapeutische Beziehung) und der Effekt spezifischer Techniken (DeRubeis et al., 2005; Siev et al., 2009). Hier steht ein allgemeiner Wirkfaktor einem spezifischen gegenüber. Welcher zählt mehr? Ohne hier auf die empirischen Befunde eingehen zu wollen, sollen dennoch ein paar Überlegungen, die auch für die vorliegende Arbeit von Interesse sind, betrachtet werden.

Einerseits sind TherapeutInneneffekt und spezifische Technik nicht immer klar voneinander zu trennen (Siev et al., 2009). Es besteht die Möglichkeit, dass manche TherapeutInnen effektiver arbeiten, weil sie die spezifischen Techniken besser anwenden oder sie dem Patienten glaubhafter vermitteln können. Es gibt auch Therapieformen, die als zentrales Element auf die TherapeutInnen-PatientInnen-Beziehung setzen. Handelt es sich dabei nicht um eine spezifische Technik? Andererseits könnte der Anteil an TherapeutInneneffekt beziehungsweise spezifischer Technik je nach Störungsbild, Behandlungsverfahren oder PatientInnengruppe unterschiedlich sein und errechnete Gesamteffekte wiederum eine Übergeneralisierung darstellen.

Diese Argumente von Siev et al. (2009) sind auch in Hinblick auf die Bona Fide-Definition von Interesse, da bei der Beurteilung non-direktiver Verfahren auch die Frage gestellt werden kann, wieviel spezifische Technik notwendig ist, um ein Verfahren als bona fide zu qualifizieren. Wampold et al. (2010) bewerteten beispielsweise ein non-direktives Verfahren als nicht-bona fide (Schnurr et al., 2007), weil sie die aktiven Bestandteile als zu gering im Vergleich zur Anwendung des Verfahrens in der Praxis einschätzten. Wenn aber in der Bona Fide-Definition ein Hinweis auf ein etabliertes Verfahren, das Vorhandensein eines Manuals, eine Beschreibung der zugrundeliegenden psychologischen Prinzipien oder die Nennung aktiver Bestandteile mit Quellennachweisen gefordert werden, bedeuten diese Kriterien nicht letztlich, dass spezifische Techniken gerade aufgrund ihres empirisch belegten oder zumindest psychologisch begründeten Wirkmechanismus eine Voraussetzung sind, um ein Therapieverfahren als bona fide zu qualifizieren? Mit anderen Worten: Werden aufgrund dieser Definition nur jene Therapieverfahren ausgewählt, deren spezifische Techniken inhaltlich wirksam sind oder deren Wirksamkeit psychologisch begründet

angenommen werden kann? Wenn ja, hätte dies Folgen für die Interpretation der empirischen Befunde über die relative Effektivität der Bona Fide-Therapien (siehe 9.2.4). Eine Differenzierung in *theoretically* und *empirically* bona fide schlugen auch Ehlers et al. (2010) in ihrer Kritik der Bona Fide-Definition vor.

Ein weiterer Aspekt der Unschärfe in den Konzepten dieser Begriffe ist, dass nach Frank (1997) das spezifische Verfahren ein allen Psychotherapien gemeinsamer allgemeiner Faktor ist. Es geht darum, dass eine spezifische Technik vorhanden ist, aber der konkrete Inhalt dieser Vorgehensweise ist nach diesem Modell nicht von Relevanz, solange sie bestimmte Funktionen erfüllt. Funktionen sind beispielsweise das Ermöglichen einer gewissen Einsicht in fehlerhafte Verhaltens- und/oder Erlebensweisen, die Gelegenheit für den Patienten, durch eigene Anstrengung die Symptome ohne Gesichtsverlust aufgeben zu können oder der Zuwachs an Selbstvertrauen durch Kompetenzentwicklung. Frank betonte ausdrücklich, dass das Aspirin, das in der Medizin-Metapher für den allgemeinen Faktor steht, ein Wirkstoff ist und kein Placebo. Dies ist wichtig festzuhalten: Die allgemeinen Wirkfaktoren sind von einer Placebowirkung abzugrenzen (Reisner, 2005), denn mitunter zielt die Kritik des Dodo Bird Verdicts auch darauf, dass es egal sei, welche Therapie man machen würde, und dies sei angesichts vorhandener empirisch validierter Therapieverfahren ein kritikwürdiges Verhalten (Carroll & Rounsaville, 2010; Chambless, 2002).

2.3.4 Was bedeutet die äquivalente Wirksamkeit von Psychotherapieverfahren?

Und letztlich der vierte Bereich: Wenn nun verschiedene Therapien gleich wirksam sind, was bedeutet das? Natürlich könnte die Äquivalenz vieler Therapieverfahren den allgemeinen Wirkfaktoren allein geschuldet sein, es wäre aber andererseits auch denkbar, dass zwei spezifische Verfahren die Symptomveränderung auf der Basis zweier unterschiedlicher Mechanismen bewirken (DeRubeis et al., 2005). Um die Medizin-Metapher weiterzuführen, wäre das eine dritte Möglichkeit, nämlich dass es durchaus unterschiedliche, aber gleich effektive Wirkprinzipien gibt.

DeRubeis et al. (2005) geben ein Beispiel aus der medikamentösen Depressionsbehandlung, aber aus der Alltagserfahrung kennt man unterschiedliche wirkungsvolle Mittel zur Behandlung von Erkältungen wie Aspirin oder Paracetamol und andere, deren Wirkung nicht nachgewiesen ist und vielleicht nicht über den Placebo-Effekt

hinausgeht, wie beispielsweise Echinacea (National Center for Complementary and Alternative Medicine, 2010). Auch in den psychotherapeutischen Interventionen könnten zusätzlich zu den allgemeinen Wirkfaktoren unterschiedliche spezifische Bestandteile für die Wirksamkeit verantwortlich sein, und zwar nicht nur, weil PatientIn und TherapeutIn ihnen eine hilfreiche Bedeutung unterstellen, sondern weil die psychologischen Mechanismen für sich genommen funktionieren. Interventionen, deren spezifische Bestandteile für ein Störungsbild aber nicht geeignet sind, wären dann jedoch eindeutig weniger wirksam. Anders als im Dodo Bird Verdict würde das bedeuten, dass nicht alle psychologischen Behandlungsverfahren gleich wirksam sind, sondern nur jene, deren spezifische Bestandteile für ein Störungsbild wirksam sind.

Aber auch hier lässt sich ein Gegenbeispiel zitieren. In einer Studie für Patientinnen mit PTSD und komorbiden substanzinduzierten Störungen (Hien, Cohen, Miele, Litt, & Capstick, 2004) richtete sich nur eines von zwei aktiven Verfahren in seinem Vorgehen an beide Störungsbilder. Dennoch gab es keine Unterschiede zwischen den Verfahren hinsichtlich ihrer Effektivität für beide Symptomatiken. Mit anderen Worten: Ein Therapieverfahren, das ursprünglich für die Behandlung eines anderen Störungsbildes gedacht war, hatte dennoch einen vergleichbaren Effekt auf die Symptome der PTSD wie das Verfahren, das spezifisch für diese Störung ausgewählt wurde.

2.4 Die Posttraumatische Belastungsstörung

Die PTSD ist eines der beiden Störungsbilder, für das eine Metaanalyse nach dem Verfahren von Wampold et al. (1997) durchgeführt wurde (Benish et al., 2008). Obwohl 2008 publiziert, umfasste diese Metaanalyse nur 15 Studien mit 17 Effekten, sodass die Autoren selbst darauf verwiesen, dass der statistische Test auf Heterogenität, den sie zur Prüfung der Nullhypothese benutzten, nach Higgins und Thompson (2002) ungefähr 20 Effekte für eine adäquate Durchführung benötigt hätte. Diese Metaanalyse rief unmittelbar Kritik hervor und löste eine Kontroverse über die Wirksamkeit traumafokussierter versus nicht-traumafokussierter Behandlungsverfahren aus (Ehlers et al., 2010; Wampold et al., 2010). Deshalb sind in der vorliegenden Arbeit Wirksamkeitsstudien zu diesem Störungsbild Gegenstand der Untersuchung, um einen aktuellen Beitrag zu dieser Diskussion zu leisten.

2.4.1 Das Störungsbild

Nach DSM-IV zählt die PTSD zur Gruppe der Angststörungen und ist eine Folgeerscheinung nach traumatischen Erlebnissen. Kennzeichnend für die PTSD ist, dass die Symptome länger als einen Monat andauern und nicht mit den unmittelbaren Reaktionen nach einem Trauma verwechselt werden dürfen. Sie ist von der akuten Belastungsstörung abzugrenzen, die im Zeitraum bis zu vier Wochen nach Erleben eines Traumas diagnostiziert wird. Nicht alle von einem Trauma Betroffenen entwickeln auch eine PTSD (Maercker, 2009a).

Obwohl die Auswirkungen traumatischer Erfahrungen historisch schon aufgrund von Folgen von Eisenbahnunglücken (Maercker, 2009a) und kriegsbedingten Phänomenen wie der *Schützengrabenneurose* bekannt waren, und Freud und Janet ähnliche Beobachtungen bei Patientinnen machten, wurde die PTSD erst 1980 von der WHO als krankheitswertige Störung anerkannt (Münker-Kramer, 2006).

Sowohl im ICD-10 als auch im DSM-IV finden sich Definitionen für den Begriff Trauma:

- „... belastendes Ereignis oder eine Situation außergewöhnlicher Bedrohung oder katastrophenartigen Ausmaßes (kurz oder langanhaltend), die bei fast jedem eine tiefe Verzweiflung hervorrufen würde“ (ICD-10; WHO, 2010, S. 183).
- „die Person erlebte, beobachtete oder war mit einem oder mehreren Ereignissen konfrontiert, die tatsächlichen oder drohenden Tod oder ernsthafte Verletzung oder eine Gefahr der körperlichen Unversehrtheit der eigenen Person oder anderen Personen beinhalteten.“ Zusätzlich zu dieser objektiven Traumadefinition wird noch eine subjektive Komponente beschrieben: „Die Reaktion der Person umfasste intensive Furcht, Hilflosigkeit oder Entsetzen“ (DSM-IV-TR; American Psychiatric Association, 2003, S. 520).

Wie Tabelle 1 zu entnehmen ist, können traumatische Ereignisse nach ihrer Verursachung und nach der Zeitkomponente unterschieden werden. Zusätzlich gibt es auch die Kategorie der medizinisch bedingten Traumata (Maercker, 2009a, S. 15).

Tabelle 1

Schematische Einteilung traumatischer Ereignisse

	Typ-I-Traumata (einmalig/ kurzfristig)	Typ-II-Traumata (mehrfach/ langfristig)	Medizinisch bedingte Traumata
Akzidentelle Traumata	Schwere Verkehrs- unfälle; Berufsbedingte Traumata (z.B. Polizei, Feuerwehr, Rettung); Kurzdauernde Kata- strophen (z.B. Wirbel- sturm, Brand)	Langdauernde Naturkata- strophen (z.B. Erdbeben, Überschwemmung); Technische Katastrophen (z.B. Giftgaskatastrophen)	Akute lebensgefähr- liche Erkrankungen (z.B. kardiale Notfälle); Chronische lebens- bedrohliche Krank- heiten (z.B. HIV/Aids); Als notwendig erlebte medizinische Eingriffe
Inter- personelle Traumata (<i>man made</i>)	Sexuelle Übergriffe (z.B. Vergewaltigung); Kriminelle bzw. körperliche Gewalt; Ziviles Gewalterleben (z.B. Banküberfall)	Sexuelle und körperliche Gewalt/Missbrauch in der Kindheit bzw. im Erwach- senenalter; Kriegserleben, Geiselhaft, Folter, politi- sche Inhaftierung	Komplizierter Behandlungsverlauf nach angenommenem Behandlungsfehler ^a

Anmerkungen. Nach "Symptomatik, Klassifikation und Epidemiologie" in A. Maercker, 2009, *Posttraumatische Belastungsstörungen* (S. 15). Heidelberg: Springer. Wiedergabe mit Genehmigung des Autors. Die Beispiele wurden etwas gekürzt.

^aDer Status dieser Eingruppierung ist noch Gegenstand wissenschaftlicher Untersuchungen.

Abgesehen vom Erleben des Traumas und der zeitlichen Dauer der Symptomatik sollten für eine Diagnose nach DSM-IV-TR Symptome aus drei Symptomgruppen vorhanden sein (Maercker, 2009a):

- **Intrusionen:** Zu dieser Gruppe gehören Symptome, die mit der Überflutung von inneren Bildern in Zusammenhang stehen: das Wiedererleben des traumatischen Ereignisses, Albträume, Flashbacks, das Erleben von Belastungen durch Auslöser, physiologische Reaktionen beim Erinnern.
- **Vermeidung/Betäubung:** Das Vermeidungsverhalten kann sowohl Gefühle, Gedanken, Aktivitäten oder Situationen betreffen. Zum Betäubungsgefühl gehören auch Amnesien, Entfremdungsgefühle, eingeschränkte Affekte, Interesselosigkeit und das Gefühl, eine eingeschränkte Zukunft zu haben.

- **Hyperarousal:** In dieser Symptomgruppe geht es um eine chronische Übererregung, die sich in Ein- und Durchschlafschwierigkeiten, erhöhter Reizbarkeit, Konzentrationsschwierigkeiten, übermäßiger Wachsamkeit und übermäßigen Schreckreaktionen äußert.

Die Lebenszeitprävalenz wurde in der *National Comorbidity Survey Replication* (Kessler et al., 2005) für die USA mit 6.8% angegeben. Für Frauen beträgt sie 9.7% und für Männern 3.6%, was auf einen deutlichen Geschlechterunterschied hinweist. Im *National Comorbidity Survey* finden sich auch Angaben über Spontanheilungen, die bei der PTSD häufig auftreten: innerhalb eines Jahres bei einem Drittel der Betroffenen und nach etwa vier Jahren bei der Hälfte der Betroffenen. 10 Jahre nach dem Trauma ist die Störung nur mehr bei einem Drittel der ursprünglich Betroffenen diagnostizierbar (Kessler et al., 1995; Maercker, 2009a).

Die häufigsten komorbiden Störungen bei der PTSD sind Angststörungen, Depressionen, Suizidalität, substanzinduzierte Störungen, Somatisierungsstörungen, Borderline- oder antisoziale Persönlichkeitsstörung und Herz-Kreislauf-Erkrankungen (Maercker, 2009a). „Bei Patienten mit PTBS [PTSD] ist das Vorliegen von Komorbidität die Regel und nicht die Ausnahme“ (Maercker, 2009b, S. 139).

Spezifische Psychotherapien für die PTSD sind: Psychodynamische Verfahren, CBT, Hypnotherapie, humanistisch-existenzielle Therapien, EMDR (*Eye Movement Desensitization and Reprocessing*) und körperbezogene Therapien (Maercker, 2009b). Die Traumafokussierung von Therapieverfahren bedeutet, in erster Linie die PTSD-Symptomatik reduzieren zu wollen, während nicht-traumafokussierte Verfahren ein allgemeineres therapeutisches Vorgehen mit den Schwerpunkten Ressourcenstärkung und Neuorientierung wählen (Maercker, 2009b). Eine andere Definition von Traumafokussierung bezieht sich darauf, dass im therapeutischen Prozess der Fokus auf die Erinnerung an das traumatische Ereignis und die individuelle Bedeutung des Traumas gelegt wird (Bisson et al., 2007; Ehlers et al., 2010). Diese Definition dient auch als Grundlage für die Kategorisierung traumafokussierter und nicht-traumafokussierter Behandlungsverfahren in der vorliegenden Arbeit.

2.4.2 Die Metaanalyse von Benish et al. (2008) und der Stand der Forschung

Die Metaanalyse von Benish et al. (2008) zur Effektivität von Bona Fide-Psychotherapien für PTSD rief unter anderem deshalb Kritik hervor (Ehlers et al., 2010), da ihre Analyse ergeben hatte, dass alle untersuchten Behandlungsverfahren gleich wirksam sind. Andere Studien lassen aber folgende Schlüsse zu: Die Therapieverfahren EMDR (Davidson & Parker, 2001) und *Prolonged Exposure* (Powers, Halpern, Ferenschak, Gillihan, & Foa, 2010) sind gleich wirksam wie andere Therapien aus dem Bereich der CBT, mit denen sie im Rahmen einer Wirksamkeitsüberprüfung verglichen wurden. In weiteren Metaanalysen zeigten sich traumafokussierte CBT (TFCBT), Expositionsverfahren und EMDR als gleich effektiv (Bisson & Andrew, 2009; Bisson et al., 2007; Bradley, Greene, Russ, Dutra, & Westen, 2005; Seidler & Wagner, 2006; Sherman, 1998; Van Etten & Taylor, 1998). Entspannungsverfahren, Hypnotherapie, Psychodynamische Therapie und Stressmanagement wiesen in diesen Studien jedoch uneinheitliche Ergebnisse auf, sodass zusammenfassend festgestellt wird, dass nicht-traumafokussierte Verfahren entweder weniger wirksam sind oder nicht genügend untersucht wurden, um ihre Wirksamkeit zu belegen. Diesbezüglich kommentierten Ehlers et al.:

Benish et al. (2008) conclude that treatments that are similar to the treatments included in their meta-analysis are equally efficacious – namely, treatments similar to “stress management, psychodynamic treatments, EMDR, hypnotherapy, cognitive behavioral treatment, exposure-based treatment, and treatments designed to explicitly exclude exposure (e.g., present-centered therapy)” (p. 755). We disagree. . . . While the present state of the PTSD literature suggests that directly addressing trauma memories in the treatment of PTSD has an advantage over non-specific factors, further research may identify active ingredients among the non-specific factors. (S. 275)

In einem Übersichtsartikel (Ponniah & Hollon, 2009), in dem PTSD-Studien nach Kriterien bewertet wurden, die in den USA verwendet werden, um neue Medikamente zuzulassen, und die eine Einteilung empirisch validierter Therapieverfahren ermöglichen, wurden aktuelle PTSD-Behandlungsverfahren zwei Kategorien zugeordnet. Als *wirksam und spezifisch* wurden jene Therapieverfahren bezeichnet, in denen die Überlegenheit gegenüber einer Placebobedingung oder einem alternativen Behandlungsverfahren in zwei oder mehreren unabhängigen Forschungskontexten

nachgewiesen wurde. Diesen Status erreichten nur die TFCBT und EMDR. Die zweite Kategorie von *möglicherweise effektiven* Verfahren, in denen die Wirksamkeit nur in einer Studie belegt werden musste, beinhaltete Stressmanagement in Kombination mit dem Stressimpfungstraining und andere Verfahren wie Hypnotherapie, Interpersonale Therapie und Psychodynamische Therapie. Kaum Evidenz gab es für Entspannungstrainings, kognitiv-verhaltenstherapeutische Gruppentherapie, Familientherapie oder supportive Beratung. Auch in den *NICE Guidelines for PTSD* (National Institute for Clinical Excellence, 2005) und den *Australian Guidelines for the treatment of adults with Acute Stress Disorder and Posttraumatic Stress Disorder* (Australian Centre for Posttraumatic Mental Health [ACPMH], 2007) wurden auf der Basis von sehr differenzierten Vergleichen zwischen verschiedenen Gruppen von Behandlungsverfahren Empfehlungen für traumafokussierte Verfahren gegeben.

Hauptkritikpunkte an der Studie von Benish et al. (2008) waren eine verzerrte Studienauswahl, da von insgesamt 27 Studien mit Direktvergleichen, die in einem Übersichtsartikel (Cloitre, 2009) angeführt wurden, nur 15 in die Metaanalyse von Benish et al. inkludiert wurden, und der fehlende Nachweis der Wirksamkeit der Verfahren (Ehlers et al., 2010). Gerade für die PTSD-Behandlungsverfahren gilt die Anforderung, dass diese wirksamer sein sollten als Spontanheilungen und die natürliche Regeneration nach einem Trauma. Außerdem besteht bei der Behandlung der PTSD die Gefahr einer Retraumatisierung und damit die Möglichkeit einer Verschlechterung der Symptomatik. *Psychological Debriefing* ist eine Krisenintervention, die kurz nach traumatischen Ereignissen durchgeführt wurde, um langfristige negative psychische Folgen für Betroffene zu vermeiden (Bisson, McFarlane, & Rose, 2000). Da empirische Befunde zeigten, dass Einzelsitzungen dieser Intervention auch keine beziehungsweise sogar negative Effekte haben können, ist es für PTSD-ForscherInnen besonders wichtig, dass Therapieverfahren für PTSD eindeutig wirksam sind. In den Inklusionskriterien der Benish et al. Metaanalyse war die nachgewiesene Wirksamkeit der Therapien aber nicht enthalten, da eine andere Zielsetzung verfolgt wurde, nämlich die Wirksamkeit therapeutisch intendierter Verfahren im Direktvergleich zu untersuchen. Außerdem kann es auch in nachgewiesenermaßen wirksamen Behandlungsverfahren wie Prolonged Exposure zu Retraumatisierungen oder unerwünschten Nebeneffekten kommen (Reisner, 2005; Sherman, 1998). Leider wird in den meisten Wirksamkeitsstudien nichts über schädliche Effekte berichtet (Bisson et al., 2007).

Die wissenschaftliche Diskussion wird auch dadurch erschwert, dass die Interpretation von Wirksamkeitsstudien alles andere als einheitlich ist. Die therapeutischen Verfahren einer Studie, die auch in der vorliegenden Arbeit enthalten ist (Brom, Kleber, & Defares, 1989), wurden von den KritikerInnen der Benish et al. (2008) Metaanalyse zum Teil als nicht wirksam bezeichnet (Ehlers et al., 2010), von den Studienautoren selbst wurde berichtet, dass sich die traumabezogenen Symptome in den Therapiegruppen im Vergleich zur Kontrollgruppe signifikant reduziert hatten, und im gerade erwähnten Übersichtsartikel basierte die Einschätzung psychodynamischer und hypnotherapeutischer Verfahren genau auf dieser Studie: „Symptoms of PTSD decreased to a significantly greater extent in patients who received psychodynamic therapy, hypnotherapy or traumafocused CBT than those who were placed on the waiting list“ (Ponniah & Hollon, 2009, S. 1103).

2.4.3 Kritik an der Bona Fide-Definition

Der Umstand, dass supportive Verfahren von Benish et al. (2008) häufig als nicht therapeutisch intendiert bezeichnet und aus der Analyse ausgeschlossen wurden, erschien den KritikerInnen als willkürlich und wurde aufgrund einiger doch in die Analyse aufgenommenen non-direktiver Therapieverfahren als nicht konsistent interpretiert (Ehlers et al., 2010). Wampold et al. (2010) betonten aber die Objektivität und Eindeutigkeit der Kriterien der Bona Fide-Definition (Kriterien siehe 4.1.1) und meinten, dass Behandlungsverfahren, die nur als Kontrollbedingung für die allgemeinen Wirkfaktoren dienen, und die meist als *Supportive Counselling* bezeichnet werden, ihrerseits zu verzerrten Ergebnissen bei Direktvergleichen aktiver Therapieverfahren führen.

In einer sehr ausführlichen Übersichtsstudie, in der TFCBT, Stressmanagement, CBT als Gruppentherapie, EMDR und eine Kategorie von *anderen Therapien* sowohl mit Kontrollbedingungen als auch miteinander verglichen wurden (Bisson & Andrew, 2007), waren in allen Vergleichen der aktiven Verfahren Studien enthalten, die im Studienkorpus von Benish et al. (2008) fehlten und damit wahrscheinlich als nicht-bona fide eingestuft wurden. Insofern ist es nachvollziehbar, dass diese Metaanalysen zu jeweils anderen Ergebnissen kamen. In einem aktuellen Übersichtsartikel zum Stand der Forschung in Bezug auf die Depressionsbehandlung stellten die AutorInnen fest, dass „Notably, whether or not the treatment is intended to be therapeutic (bona fide) is likely a determining factor in questions of effectiveness“ (Berg & Høie,

2010, S. 2199). Auch in einer der aktuellsten Metaanalysen zur relativen Effektivität von Psychotherapien wurden nur Bona Fide-Therapien inkludiert (Tolin, 2010).

Kritisiert wurde an der Definition aber auch, dass ein Behandlungsverfahren an sich bona fide sein kann, ohne dabei zu präzisieren für welches Störungsbild. Außerdem veränderten sich manche etablierte Verfahren bezüglich ihrer Anwendbarkeit im Laufe der Zeit, sodass sich auch in dieser Beziehung die Frage stellt, ob diese nun bona fide sind oder nicht (Siev et al., 2009). Entspannungstechniken beispielsweise, die früher zur Behandlung von Zwangsstörungen eingesetzt wurden, sind heute für dieses Störungsbild keine adäquate Therapieform mehr, aber für die generalisierte Angststörung noch immer das Verfahren der Wahl (ebd.).

Dieser Gesichtspunkt wird besonders auffällig, wenn man Einzelstudien aus der Metaanalyse der alkoholinduzierten Störungen betrachtet (Imel et al., 2008). In diese Analyse wurden einige Studien integriert und damit Therapieverfahren als bona fide qualifiziert, die heute nicht mehr zeitgemäß sind. Als Beispiel soll hier ein Verfahren aus der Studie von Hedberg und Campbell (1974) zitiert werden, das sich abgesehen von einer hohen Ausstiegsrate auch als nicht wirksam erwiesen hat:

Faradic Stimulation: Twelve patients were initially assigned to the faradic stimulation treatment condition, but only four remained in the treatment program beyond the third session. The treatment procedure followed the avoidance conditioning model set forth by Feldman and MacCullough (1965). Each patient selected 15 alcohol stimulus slides and 15 relief slides from a library of slides. All slides were selected on the basis of their relevance to the patients own drinking pattern. Each session consisted of 30 alcohol related slide presentations in which contingent faradic shock was administered to the patient through either finger or forearm electrodes. A portable shock generator, powered by a nine volt battery was used to produce the electric stimulus administered to the patients. (Hedberg, 1974, presentation copy)

Obwohl die Kritik an der Bona Fide-Definition beziehungsweise an der Vollständigkeit der Kriterien durchaus berechtigt ist, bleibt der Umstand bestehen, aufgrund dieser Definition über ein Trennkriterium zwischen therapeutisch intendierten Behandlungsverfahren und Kontrollbedingungen zu verfügen.

3 Zielsetzung und Fragestellung

Die zentrale Frage dieser Diplomarbeit lautet: Sind Bona Fide-Psychotherapien für die PTSD gleich effektiv? Das Ziel war es, dieser Frage nach den Unterschieden in der Wirksamkeit von psychologischen Therapieverfahren differenziert nachzugehen:

1. Welche Rolle spielt die metaanalytische Methodik? Sind äquivalente Ergebnisse unter Umständen Artefakte einer spezifischen Methodik? Das Verfahren, das in den Metaanalysen von den Forschergruppen um Wampold verwendet wurde, wurde repliziert und um aktuelle Studien ergänzt nochmals durchgeführt. Dabei wurde eine möglichst differenzierende Strategie bei den untersuchten Effekten verwendet, anstatt über alle Maße zu aggregieren, um allfällige Unterschiede auch tatsächlich sichtbar zu machen.

2. Sind Kategorien von Behandlungsverfahren signifikant wirksamer als Vergleichsverfahren? Welche Therapierichtungen zeigen äquivalente Ergebnisse, und wo sind Unterschiede in der Effektivität zu finden? Diese Frage sollte auch in Hinblick auf therapeutisch intendierte supportive Verfahren beantwortet werden, da diese Kategorie bis jetzt noch nie in dieser Form auf ihre Wirksamkeit zur Behandlung der PTSD untersucht wurde. Darauf aufbauend sollte die Frage beantwortet werden, ob traumafokussierte Behandlungsverfahren effektiver sind als nicht-traumafokussierte Verfahren, wenn ebenfalls sichergestellt ist, dass letztere therapeutisch intendiert und keine psychologischen Placebos sind. Alle verfügbaren PTSD-Wirksamkeitsstudien mit Direktvergleichen wurden Kategorien zugeordnet, und für jede wurden möglichst differenziert Metaanalysen durchgeführt, um einen aktuellen Beitrag zum Stand der Forschung in Bezug auf die relative Effektivität von PTSD-Behandlungsverfahren leisten zu können.

3. Welche Rolle spielt das Störungsbild? Könnte es sein, wie manche Studien nahelegen (Siev & Chambless, 2007; Tolin, 2010), dass sich bei einem Störungsbild die psychologischen Therapien in ihrer Wirksamkeit unterscheiden und bei einem anderen Störungsbild nicht? Die vorliegende Arbeit kann zur Beantwortung dieser Frage einen Teilaspekt einbringen. Eine weitere Metaanalyse zu dieser Frage wird derzeit von Sarah Braun (2012) durchgeführt.

4 Studienauswahl und Suchstrategie

Untersuchungsgegenstand der vorliegenden Metaanalysen sind randomisierte, kontrollierte Primärstudien, die über einen direkten Vergleich mindestens zweier Bona Fide-Psychotherapien zur Behandlung der PTSD berichten. Die Metaanalyse von Benish et al. (2008) umfasste 15 dieser Studien, die zwischen 1989 und 2005 publiziert wurden. Da in der vorliegenden Arbeit diese Metaanalyse repliziert und aktualisiert wurde, wurden die Einschluss- und Ausschlusskriterien von Benish et al. für die Suche nach ergänzenden Studien exakt übernommen, und es wurde eine weitgehend deckungsgleiche Suchstrategie durchgeführt.

4.1 Inklusions- und Exklusionskriterien

Inklusionskriterien auf der Designebene waren ein randomisiertes, kontrolliertes Design und der Vergleich mindestens zweier Behandlungsverfahren. Diese mussten der Bona Fide-Definition entsprechen und mindestens zwei oder mehr Therapiesitzungen umfassen. StudienteilnehmerInnen waren Erwachsene mit einer PTSD-Diagnose nach DSM-III oder DSM-IV, deren PTSD-Symptomatik erhoben wurde. Darüber hinaus musste es möglich sein, die Effektgrößen auf der Basis der publizierten statistischen Daten zu berechnen. Ausgeschlossen wurden alle Varianten von Komponentenstudien, in denen sich die Behandlungsverfahren nur in Bezug auf das Vorhandensein beziehungsweise das Ausmaß eines bestimmten Bestandteils des Verfahrens unterschieden. Ausgeschlossen wurden auch Artikel, in denen nur Zusatzanalysen zu einer bereits veröffentlichten Studie berichtet wurden. Damit wurde sichergestellt, dass jeder Datensatz nur einmal in die Analyse einging.

4.1.1 Die Bona Fide-Definition

Um ein Behandlungsverfahren als bona fide zu qualifizieren, mussten zwei von den folgenden vier Kriterien erfüllt sein (Benish et al., 2008; Wampold et al., 2010):

- a. Es wurde ein etabliertes psychologisches Verfahren zitiert (z.B. Klientenzentrierte Psychotherapie nach Rogers, 1973).
- b. Das Behandlungsverfahren wurde beschrieben und basierte auf psychologischen Prinzipien (z.B. operantes Konditionieren).

- c. Es war ein Manual des Behandlungsverfahrens vorhanden, das als Leitfaden für die Therapie verwendet wurde.
- d. Es wurden aktive Bestandteile des Behandlungsverfahrens mit Quellenangabe benannt.

Zusätzlich musste das Verfahren von ausgebildeten TherapeutInnen durchgeführt werden, die es an die PatientInnen anpassten, und Kommunikationselemente enthalten, die den Aufbau einer TherapeutInnen-PatientInnen-Beziehung ermöglichen. Das bedeutet, dass ausschließlich medienvermittelte Behandlungsverfahren oder Verfahren, die es nicht erlaubten, auf die Reaktionen des Patienten einzugehen, die Bona Fide-Definition nicht erfüllten (z.B. Audiobänder mit Programmen zur Progressiven Muskelentspannung).

Um die Bona Fide-Definition konsistent als Einschlusskriterium für neue Studien anwenden zu können, wurden als Vorbereitung die Beschreibungen von etwa 112 als bona fide qualifizierten Behandlungsverfahren dreier Störungsbilder (PTSD: Benish et al., 2008; Alkoholinduzierte Störungen: Imel et al., 2008; Depression: Wampold et al., 2002) gelesen, sowie die Beschreibungen von jenen Behandlungsverfahren, die in diesen Studien als nicht-bona fide ausgeschlossen wurden. Besonderes Augenmerk wurde auch auf die Argumentation gelegt, die Wampold et al. (2010) zur Klärung der Bona Fide-Definition ausführlich an vier Beispielen darlegten.

Da in Primärstudien mitunter auch mehr als zwei Therapieverfahren miteinander verglichen wurden, war es wesentlich, sowohl die Einschlusskriterien auf jedes einzelne Verfahren anzuwenden, als auch das Ausschlusskriterium der Komponentenstudie innerhalb der Behandlungsverfahren einer Studie zu überprüfen. Jene Verfahren, die nicht als bona fide qualifiziert wurden, oder die Teilaspekte beziehungsweise Erweiterungen anderer in der Studie inkludierter Verfahren darstellten, wurden von der Analyse ausgeschlossen. Das Ergebnis der Studienausswahl war deshalb sowohl eine bestimmte Anzahl von Primärstudien als auch eine bestimmte Anzahl von direkten Vergleichen.

4.2 Suchstrategie

Da die aktuellsten Primärstudien in der Metaanalyse von Benish et al. (2008) aus dem Jahr 2005 stammten, wurde die Suche nach neuen Studien auf die Jahre 2005 bis März 2011 beschränkt.

Eine effektive Suchstrategie beruht auf mehreren Suchpfaden, wobei folgende Quellen empfohlen werden (Lipsey & Wilson, 2001):

- Übersichtsartikel, Metaanalysen und Literaturverzeichnisse von Primärstudien zum relevanten Thema.
- Computerdatenbanken: Es wird empfohlen, die Suche in unterschiedlichen Datenbanken durchzuführen und dabei auch Datenbanken zu berücksichtigen, die nicht publizierte Literatur wie zum Beispiel Dissertationen zugänglich machen (z.B. *Dissertations Abstracts Online*).
- Händisch durchgeführte Suche in einschlägigen Fachzeitschriften und bibliografischen Jahrbüchern.
- Direkter Kontakt zu WissenschaftlerInnen, ExpertInnen, Regierungsstellen und professionellen Organisationen, die sich mit dem jeweiligen Thema befassen und Forschungsberichte publizieren oder Konferenzen durchführen.

Benish et al. (2008) suchten in den Datenbanken MEDLINE, Cinahl Health, Healthsource, PsycINFO, PsycArticles, PubMed, Social Sciences Fulltext, Web of Knowledge mit den Suchbegriffen *PTSD*, *posttraumatic*, *post-traumatic*, *post traumatic*, *psychotherapy* und *treatment* und überprüften die Literaturverzeichnisse von Metaanalysen, Primärstudien und Übersichtsartikel zu PTSD Behandlungsverfahren. Es wurden keine Forschungsarbeiten aus nicht publizierter Literatur in die Analyse von Benish et al. aufgenommen.

Damit die Stichprobe der Studien von 2005 bis 2011 in ihrer Zusammensetzung möglichst mit Benish et al. (2008) übereinstimmte, wurde die Suchstrategie weitgehend übernommen, obwohl damit nicht alle Quellen ausgeschöpft wurden und sie somit nicht zur Gänze dem momentanen Standard für Metaanalysen entspricht.

Folgendes ist außerdem festzuhalten:

- Statt der Datenbank Healthsource wurde die Datenbank *Psychology and Behavioral Sciences Collection* benutzt.
- Zusätzlich zu den oben genannten Suchbegriffen wurden auch *randomized controlled trial*, *relative efficacy*, *outcome study* und *comparative efficacy* als ergänzende Suchbegriffe verwendet.
- Aktuelle Übersichtsartikel (ACPMH, 2007; Ponniah & Hollon, 2009) und Metaanalysen (Bisson et al., 2007; Powers et al., 2010; Seidler & Wagner, 2006; Taylor & Harvey, 2009) wurden auf neue Primärstudien hin gelesen und die Literaturverzeichnisse überprüft.
- Da Ehlers et al. (2010) darauf hinwiesen, dass Cloitre (2009) für einen ähnlichen Suchzeitraum 27 Studien mit 44 Direktvergleichen im Gegensatz zu den 15 Studien mit 17 Vergleichen von Benish et al. (2008) identifiziert hatten, wurden jene 27 Studien nochmals gesondert hinsichtlich der Einschluss- und Ausschlusskriterien überprüft.

Die Vorgehensweise zur Identifizierung neuer Primärstudien in den Datenbanken bestand aus drei Arbeitsschritten. Zuerst wurden die Titel von allen Treffern nach Eingabe der Suchbegriffe gelesen und dabei jene aussortiert, die nicht dem Untersuchungsgegenstand entsprachen. Von den restlichen Treffern wurden die Kurzzusammenfassungen gelesen, um festzustellen, ob Design, Zielgruppe und Störungsbild die Kriterien erfüllten. Von den in Frage kommenden Studien wurden dann die Originalartikel gelesen, um die Behandlungsverfahren hinsichtlich der Bona Fide-Definition zu überprüfen. Bei nicht eindeutig zu treffenden Entscheidungen und methodischen Fragen hinsichtlich des Designs (z.B. bei Komponentenstudien) wurde Rücksprache mit dem Betreuer dieser Arbeit gehalten (Doktor der Psychologie, Klinischer Psychologe, Lektor und wissenschaftlicher Mitarbeiter der Fakultät für Psychologie an der Universität Wien).

5 Kodierung

Sobald der Korpus von Primärstudien zusammengestellt war, konnte mit der Kodierung begonnen werden. Die Kodierung ähnelt der Umfrageforschung, bei der anhand eines Fragebogens Informationen erhoben werden. Der Fragebogen entspricht in der Metaanalyse dem Kodierprotokoll, anhand dessen die Informationen aus den Primärstudien extrahiert werden (Lipsey & Wilson, 2001). Den Empfehlungen von Lipsey und Wilson folgend, wurde für die vorliegenden Metaanalysen ein Kodierprotokoll erstellt, das vier Bereiche mit folgenden Informationen umfasste:

- **Studiendurchführung:** Publikationsjahr, Art der Publikation, Vorhandensein einer Kontrollgruppe, Durchführungsland, Informationen zum Studiendesign zur Bewertung der Studienqualität (siehe 5.4.1).
- **PatientInnenstichprobe:** Anzahl der PatientInnen, Geschlechterzusammensetzung der Stichprobe, Durchschnittsalter, Art des Traumas, Zeitdauer seit Bestehen der Traumatisierung, Schweregrad der PTSD, Komorbiditäten (Depression, Angst, Alkohol- und Substanzmissbrauch), Rekrutierung der Stichprobe (öffentliches oder klinisches Umfeld), Psychopharmakagebrauch.
- **Behandlungsverfahren:** Zuordnung der Behandlungsverfahren zu Kategorien, Individual- oder Gruppenverfahren, Länge der Behandlung in Wochen, Anzahl und Dauer der Sitzungen.
- **Messeigenschaften:** Zuordnung der Messinstrumente zu den Variablen, Abbrecherquoten pro Behandlungsverfahren, Anzahl der PatientInnen ohne Diagnose nach der Behandlung, Nacherhebungen (*Follow-ups*), Daten zur Berechnung der Effektgrößen (wenn möglich Mittelwerte und Standardabweichungen).

Das Kodierprotokoll wurde mit dem Betreuer dieser Arbeit diskutiert und abgestimmt, um so weit wie möglich alle methodischen Entscheidungen vor dem Kodierprozess zu treffen. Die Kodierung wurde dann von der Verfasserin dieser Arbeit vorgenommen. Da aufgrund der uneinheitlichen und lückenhaften Berichterstattung wissenschaftlicher Studien immer wieder begründete Annahmen getroffen

werden müssen (Lipsey & Wilson, 2001), wurden alle auftretenden Unstimmigkeiten im Kodierprozess mit dem Betreuer besprochen und geklärt. Um die Reliabilität der Kodierung zu erhöhen, könnte man nach Lipsey und Wilson bei nicht so umfangreichen Metaanalysen zwei Kodierer einsetzen, die in regelmäßigen Abständen ihre Kodierprotokolle diskutieren und vergleichen, was aber in der vorliegenden Arbeit nicht möglich war. Allerdings wurden stichprobenartig Vergleiche mit anderen Metaanalysen vorgenommen, wenn dort die gleichen Studien in den Analysen enthalten waren (ACPMH, 2007; Bradley et al., 2005; Powers et al., 2010; Seidler & Wagner, 2006; Taylor & Harvey, 2009). Die methodischen Entscheidungen zum Kodierprotokoll werden nun zusammengefasst.

5.1 PatientInnenanzahl

Primärstudien berichten ihre Untersuchungsergebnisse entweder für die *Intent-to-treat* (ITT-)Stichprobe und/oder für jene PatientInnen, die die Behandlung abgeschlossen haben (*Completer*). Als ITT-Stichprobe werden jene PatientInnen definiert, die nach der Randomisierung einem Behandlungsverfahren zugeteilt wurden oder manchmal auch jene PatientInnen, die noch zusätzlich an mindestens einer therapeutischen Sitzung teilgenommen haben. Es gibt auch keine einheitliche Vorgehensweise, welche Messwerte der ITT-Stichprobe die StudienautorInnen benutzen. Manchmal werden die Werte verwendet, die unmittelbar vor dem Therapieabbruch erhoben wurden, manchmal werden aber auch jene Werte herangezogen, die von der Erhebung zu Beginn des Behandlungsverfahrens stammen, unabhängig davon, wann die Therapie abgebrochen wurde. Als Completer gelten jene PatientInnen, die die Behandlung abgeschlossen haben. In manchen Studien werden auch jene PatientInnen als Completer bezeichnet, die eine zuvor definierte Anzahl von Sitzungen absolviert haben.

Das Verwenden der Messwerte der ITT-Stichprobe in einer Metaanalyse führt zu konservativeren Ergebnissen, da hier die Annahme zugrunde liegt, dass das Behandlungsverfahren für die TherapieabbrecherInnen nicht erfolgreich oder wirksam war. Darüber hinaus bieten nur ITT-Analysen durch die Randomisierung einen Schutz vor systematischer Verzerrung (Staines, 2008). Deshalb wurden für die vorliegende Arbeit, wenn vorhanden, die Messwerte der ITT-Stichproben verwendet, und nur falls diese nicht vorhanden waren, die Messwerte der Completer-Stichproben.

5.2 Variablen

Benish et al. (2008) berechneten für zwei verschiedene Variablen Effektgrößen und führten damit je eine Metaanalyse durch. Einmal fassten sie unter dem Begriff *PTSD Symptoms only* all jene Maße zu einer Effektgröße zusammen, die sich direkt auf die PTSD bezogen und die größtenteils mit folgenden Messinstrumenten erhoben wurden: Clinician Administered PTSD Scale (CAPS), Impact of Events Scale (IES), State-Trait Anxiety Inventory (STAI) und PTSD Symptom Scale (PSS-I bzw. PSS-SR). Zum anderen fassten sie alle Ergebnisse (*all post-treatment measures of psychological functioning*) mit dem Hinweis auf die außergewöhnlich hohen Komorbiditätsraten in der PTSD zu einer zweiten Effektgröße zusammen. Diese stellte damit eine Erweiterung der ersten Variable dar.

Eine differenzierende Erfolgsmessung für psychologische Interventionen sollte, wie unter 2.3.2 dargelegt, auf mehreren Variablen beruhen. Es liegt auch in der Tradition der Metaanalysen zur Wirksamkeit von PTSD-Behandlungsverfahren, dass die Effektivität über unterschiedliche Variablen abgebildet und verglichen wird. Bisson et al. (2007) beispielsweise untersuchten folgende Variablen: als kontinuierliche Variablen die PTSD-Symptomatik getrennt nach Selbstbeurteilungsfragebögen und Interviews, Maße zu komorbider Angst und Depression, sowie die dichotomen Variablen PTSD-Diagnose und Therapieabbruchrate. Powers et al. (2010) wiederum unterschieden zwischen *PTSD symptom severity* und *General subjective distress*, wobei erstere nur Maße zur PTSD und zweitere Maße zur Depression, Angst, gesundheitsbezogenen Lebensqualität und zum sozialen Funktionsniveau enthielten.

Um die Forschungsfrage nach den Unterschieden zwischen den Behandlungsverfahren differenziert beantworten zu können, wurden in der vorliegenden Arbeit die folgenden Variablen definiert.

5.2.1 Primäre Maße

Ein zentrales Forschungsinteresse eines Wirksamkeitsnachweises ist es, zu erheben, wie sich das Therapieverfahren auf das Störungsbild auswirkt (Siev et al., 2009). Primäre Maße sollten daher störungsspezifisch sein. Deshalb wurden als primäre Maße jene angesehen, wo explizit Symptome der PTSD erfasst wurden. Am Beispiel des STAI lässt sich die unterschiedliche Zuordnung im Gegensatz zu Benish et al. (2008) ver-

deutlichen. Das STAI misst als Angstfragebogen zwar ein wesentliches Symptom der PTSD, der Fragebogen ist jedoch nicht explizit auf die PTSD bezogen. Obwohl die PTSD zu den Angsterkrankungen gehört, richten sich die störungsspezifischen Messverfahren explizit an die drei Symptomgruppen des Störungsbildes. Entsprechend seiner Konzeption, Angst als momentanen Zustand (*State*) und Angst als Eigenschaft (*Trait*) zu messen, wurde der Fragebogen in den Primärstudien außerdem unterschiedlich eingesetzt. Größtenteils wurde nur die State-Angstskala erhoben, fallweise auch beide Skalen (Brom et al., 1989; Paunovic & Öst, 2001; Rothbaum, Astin, & Marsteller, 2005) beziehungsweise nur die Trait-Angstskala (Deville & Spence, 1999). Angst als Eigenschaft ist jedoch ein Persönlichkeitskonstrukt, das nicht direkt in Bezug zur Symptomatik der PTSD steht. Außerdem erhält man mit dem STAI nur eine allgemeine und eher ungenaue Zustandsbeschreibung einer PatientInnengruppe (Kubinger, 2009).

Würde man alle Angstfragebögen zu den primären Maßen zählen wie Benish et al. (2008), vermischt sich das Störungsbild der PTSD mit einer, aber eben nur einer häufigen komorbiden Störung, nämlich anderen Angsterkrankungen. 80% der PatientInnen mit einer Lebenszeit-PTSD-Diagnose leiden unter zumindest einer der drei häufigsten komorbiden Störungen bei PTSD (Foa, Keane, & Friedman, 2000), wie auch Bradley et al. (2005) festhalten: „The most common comorbid conditions include depression, substance abuse, and other anxiety disorders“ (S. 214). Der spezifische Effekt eines Behandlungsverfahrens auf die PTSD kann daher mit der Variable *PTSD Symptoms only*, so wie Benish et al. sie definiert haben, nur ungenau dargestellt werden. Deshalb macht es Sinn, die Effekte der Behandlungsverfahren auf komorbide Störungen getrennt zu erheben, sodass in dieser Arbeit das STAI ebenso wie andere allgemeine Angstfragebögen den sekundären Maßen zugeordnet wurden. Dies entspricht auch dem Vorgehen einer neueren Metaanalyse zu Behandlungsverfahren der PTSD (Powers et al., 2010). Aufgrund der unterschiedlichen Definitionen der Primärmaße von Benish et al. und dieser Arbeit unterscheiden sich teilweise auch die Effektgrößen, die in die jeweiligen Analysen eingingen (siehe 7.3).

Dem Beispiel von Bisson et al. (2007) folgend, wurden die primären Maße zusätzlich getrennt nach Selbstbeurteilungsfragebögen und Interviews erfasst, da die metaanalytischen Ergebnisse zu diesen beiden Erhebungsmethoden nicht einheitlich sind. Van Etten und Taylor (1998) konnten keine signifikanten Unterschiede

zwischen den Effektgrößen je nach Erhebungsmethode feststellen, aber in anderen Metaanalysen gab es doch Hinweise auf Unterschiede (Cuijpers, Li, Hofmann, & Andersson, 2010; Taylor & Harvey, 2009). Die Zuordnung der Messinstrumente zu den primären Maßen ist Tabelle 2 zu entnehmen.

Tabelle 2

Zuordnung der Messinstrumente zur Variable Primäre Maße

PTSD Interviews		Studien ^a
CAPS	Clinician Administered PTSD Scale (Blake et al., 1995)	14
PSS-I	PTSD Symptom Scale (Foa, Riggs, Dancu, & Rothbaum, 1993)	2
PTSD-I	PTSD Interview (Watson, Juba, Manifold, Kucala, & Anderson, 1991)	1
SI-PTSD	Davidson's Structured Interview for PTSD (Davidson, Smith, & Kudler, 1989)	1
CGI	Clinical Global Impression of PTSD (Guy, 1976)	1
PTSD Selbstbeurteilungsfragebögen		
IES/IES-R	Impact of Events Scale/Impact of Events Scale - Revidierte Form (Horowitz, Wilner, & Alvarez, 1979)	9
PSS-SR	PTSD Symptom Scale Self-Report (Foa et al., 1993)	5
PCL	PTSD Checklist (Weathers, Litz, Herman, Huska, & Keane, 1993)	4
PENN	Penn Inventory (Hammarberg, 1992)	1
SI-PTSD	Davidson's Structured Interview for PTSD, self-report version (Davidson et al., 1989)	1
PDS	Post-traumatic Stress Diagnostic Scale (Foa, 1995)	1
CMS	Civilian Mississippi Scale for PTSD (Keane, Caddell, & Taylor, 1988)	1
MMPI-K	Keane's Post-Traumatic Stress Disorder Scale from the Minnesota Multiphasic Personality Inventory (Keane, Malloy, & Fairbank, 1984)	1

Anmerkungen. Die Messinstrumente wurden hier nur in der Originalversion angeführt.

^aAnzahl der Studien, in denen dieses Messinstrument verwendet wurde.

5.2.2 Sekundäre Maße

Als sekundäre Maße wurden die Maße der häufigsten komorbiden Störungen, traumabezogene Symptome und Konstrukte und Maße zur allgemeinen Symptombelastung und zum sozialen Funktionsniveau zusammengefasst, ohne jedoch die primären Maße wieder einzubeziehen. In Anlehnung an Benish et al. (2008) könnte man diese Variable als *all other measures of psychological functioning* bezeichnen. Im

Sinne der Fragestellung, ob alle Bona Fide-Psychotherapien gleich wirksam sind, könnten sich Behandlungseffekte in der komorbiden Symptomatik oder im sozialen Funktionsniveau zeigen, die unter Umständen verborgen blieben, würde man alle Maße aggregieren. Um auch diese Variable differenziert betrachten zu können, wurden alle mit der PTSD in Beziehung stehenden störungsbezogenen Symptome und Konstrukte auch getrennt von den funktionsbezogenen Maßen wie der gesundheitsbezogenen Lebensqualität und dem sozialen Funktionsniveau erfasst. Die Zuordnung der Messinstrumente zu den sekundären Maßen ist Tabelle 3 zu entnehmen.

Tabelle 3

Zuordnung der Messinstrumente zur Variable Sekundäre Maße

Depression, Angst, Substanz- und Alkoholkonsum, traumabezogene Symptome und Konstrukte wie Schuld, Ärger, Dissoziation und entsprechende Kognitionen		Studien ^a
BDI	Beck Depression Inventory (Beck, Ward, Mendelsohn, Mock, & Erbaugh, 1961)	16
HAM-D	Hamilton Rating Scale for Depression (Hamilton, 1960)	1
MADRS	Montgomery Asberg Depression Rating Scale (Montgomery & Asberg, 1979)	1
HADS	The Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983)	1
SRQ-20	Self-Reporting Questionnaire 20 (Harding et al., 1980)	1
STAI	State-Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1970)	9
BAI	Beck Anxiety Inventory (Beck, Epstein, Brown, & Steer, 1988)	2
FQ	Fear Questionnaire (Marks & Mathews, 1979)	1
HAM-A	Hamilton Rating Scale for Anxiety (Hamilton, 1959)	3
ASI	Addiction Severity Index (McLellan et al., 1992)	1
SUI	Substance Use Inventory (Sobell et al., 1980)	1
RAST	Rape Aftermath Symptom Test (Kilpatrick, 1988)	1
NFQ	Night Frequency Questionnaire (Krakow et al., 2000)	1
PSQI	Pittsburgh Sleep Quality Index (inkl. Addendum for PTSD) (Germain, Hall, Krakow, Shear, & Buysse, 2005)	1
NES	Nightmare Effects Survey (Krakow et al., 2000)	1
TRGI	Trauma-Related Guilt Inventory (Kubany et al., 1996)	1
COOK	Cook-Medley Hostility Scale (Cook & Medley, 1954)	1
STAXI	State-Trait Anger Expression Inventory (Spielberger, 1988)	2

DES (-II)	Dissociative Experiences Scale(-II) (Bernstein & Putnam, 1986)	2
TSI	Traumatic Stress Institute Beliefs Scale (Pearlman, 2001)	1
CCQ	Catastrophic Cognitions Questionnaire (Khawaja & Oei, 1992)	1
WAS	World Assumptions Scale (Janoff-Bulman, 1989)	1
Gesundheitsbezogene Lebensqualität, Gesundheit und soziales Funktionsniveau		
QOLI	Quality of Life Inventory (Frisch, Cornell, Villanueva, & Retzlaff, 1992)	3
QOLS	Quality of Life Scale (Marks et al., 1993)	1
GHQ 28	General health questionnaire (Goldberg & Hillier, 1979)	2
SF-36/12	Short-Form Health Survey (Ware & Sherbourne, 1992)	4
SCL-90 (R)	Symptom Checklist-90 (Arindell & Ettema, 1981)	2
SHEEHAN	The Sheehan Disability Scale (Sheehan, 1986)	1
SAS	Social Adjustment Scale (Weissman & Paykel, 1974)	1

Anmerkungen. Die Messinstrumente wurden hier nur in der Originalversion angeführt.

^aAnzahl der Studien, in denen dieses Messinstrument verwendet wurde.

5.2.3 Follow-ups

Das Bewerten der Wirksamkeit von Behandlungsverfahren über einen längeren Zeitraum hinweg wäre auch aus metaanalytischer Sicht wesentlich. Behandlungsverfahren könnten kurz nach Therapieende bedeutende Effekte zeigen, aber ob diese Wirkung anhält, wäre ein wesentliches Kriterium zur Unterscheidung von wirksamen und weniger wirksamen Verfahren (Bradley et al., 2005). Allerdings sind bis jetzt kaum Effektgrößen von direkten Vergleichen von PTSD-Behandlungsverfahren zu Follow-up-Zeitpunkten metaanalytisch untersucht worden. Die wenigen Befunde legen nahe, dass kaum Unterschiede zwischen den untersuchten Verfahren dieser Studien vorhanden sind (Bradley et al., 2005; Van Etten & Taylor, 2009). Benish et al. (2008) verzichteten auf eine Analyse der Effektgrößen zu einem Follow-up-Zeitpunkt mit dem Argument, dass nicht genügend Daten vorlagen.

Für diese Arbeit wurden die primären und sekundären Maße sowohl zum Therapieende als auch zum ersten und – so vorhanden – zum zweiten Follow-up analysiert, um der Forschungsfrage nach Unterschieden zwischen den Behandlungsverfahren zumindest zu einem Follow-up-Zeitpunkt nachgehen zu können. Obwohl Bradley et al. (2005) vorschlugen, Nacherhebungsdaten erst ab sechs Monaten zu erfassen, wurde in dieser Arbeit kein zeitliches Limit gesetzt.

5.2.4 Klinische Signifikanz

Die Wirksamkeit zweier Behandlungsverfahren kann durch den Vergleich der primären und sekundären Maße dieser Verfahren beschrieben werden. Da es jedoch vorstellbar ist, dass ein Behandlungsverfahren eine deutliche Besserung der Symptomatik bewirkt, und die psychische Störung dennoch weiterhin diagnostiziert werden kann, stellt sich die Frage nach der klinischen Bedeutsamkeit des Effekts: Wie hoch ist der Anteil der PatientInnen, die erfolgreich behandelt wurden?

In Wirksamkeitsstudien werden unterschiedliche Kriterien für *erfolgreich* verwendet. Bradley et al. (2005), die diese Frage sehr differenziert betrachten, erweiterten ihre Metaanalyse um zwei zusätzliche Indizes: die Genesungsrate und die Besserungsrate. Die Genesungsrate bezieht sich auf die Anzahl der PatientInnen, die nach Abschluss der Therapie nicht mehr mit PTSD diagnostiziert werden können, die Besserungsrate wird je nach StudienautorInnen unterschiedlich definiert. Häufig bedeutet eine Besserung eine Abnahme in den Ergebnissen des Messinstruments um zwei oder mehr Standardabweichungen. Bradley et al. analysierten beide Indizes sowohl für die ITT- als auch für die Completer-Stichprobe, da sie auf diese Weise einen liberalen und einen konservativen Wert der Indizes erhielten. Außerdem erhoben sie einen *Posttreatment Symptom Level*, um noch einen weiteren Indikator für die klinisch relevante Veränderung zu haben.

Für die vorliegende Arbeit wurde ein weniger detailliertes Vorgehen gewählt, da die Annahme bestand, dass die für eine differenzierte Analyse notwendigen Daten in zu wenigen Studien berichtet worden waren. Vielmehr wurde für jede Studie eine Variable *Klinische Signifikanz* kodiert, wobei es sich in erster Linie um die Genesungsrate handeln sollte. Für den Fall dass keine Informationen zur Genesungsrate vorlagen, wurde die Besserungsrate verwendet. Westen und Bradley (2005) betonten, dass die Entscheidung, auf Basis welcher Personenanzahl diese Indizes berechnet werden, wesentlich ist. Um die konservativste Vorgehensweise zu wählen, die die vorhandenen Daten erlaubten, wurde in der vorliegenden Arbeit die Kodierung analog zu den primären und sekundären Maßen vorgenommen. Es wurden die Daten der ITT-Stichproben herangezogen, wenn diese von den StudienautorInnen berichtet wurden. Ansonsten wurden die Daten der Completer-Stichproben verwendet.

5.2.5 Dropouts

PatientInnen, die eine Therapie abbrechen, werden auch als *Dropouts* bezeichnet. Da einzelne Primärstudien deutliche Unterschiede in der Anzahl der Personen berichteten, die eine Intervention im Vergleich zu einem anderen aktiven Verfahren abgebrochen hatten (z.B. Cottraux et al., 2008; Ironson, Freund, Strauss, & Williams, 2002; McDonagh et al., 2005; Schnurr et al., 2003), wurde in dieser Arbeit überprüft, ob es tatsächlich einen Unterschied hinsichtlich der Abbruchraten gibt. Dies ist vor allem aus klinischer Sicht von Bedeutung, wenn sich die Therapien in ihren primären und sekundären Maßen nur geringfügig unterscheiden. Welches Verfahren es mehr PatientInnen ermöglichen würde, ihre Therapie abzuschließen, wäre dann für die Erfolgsbewertung ausschlaggebend. In den bisher durchgeführten Metaanalysen zu PTSD-Behandlungsverfahren wurden eher geringe Unterschiede in den Abbruchraten berichtet (Bisson et al., 2007; Bradley et al., 2005; Van Etten & Taylor, 1998).

5.3 Kategorisierung der Behandlungsverfahren

Während in einer der ersten Metaanalysen zum Vergleich von PTSD-Psychotherapien (Van Etten & Taylor, 1998) die Kategorien der aktiven psychologischen Interventionen zum Teil noch sehr breit waren (Verhaltenstherapie, EMDR, Entspannungsverfahren, andere Verfahren), wurden sie mit der Zeit spezifischer, und schließlich wurden eigene Metaanalysen für einzelne Therapieformen der PTSD durchgeführt (Davidson & Parker, 2001; Powers et al., 2010; Seidler & Wagner, 2006). Die Zuordnung der Interventionen zu den von den StudienautorInnen festgesetzten Kategorien wurde nicht einheitlich gehandhabt. Ein sehr aufwendiges und fundiertes Kategorisierungsverfahren für eine vergleichende Metaanalyse im Bereich Depression basierte auf eindeutigen, von ExpertInnen im Konsens erarbeiteten Definitionen der Verfahren (Cuijpers et al., 2008).

In dieser Arbeit wurde die Kategorisierung wie folgt durchgeführt: In einem ersten Schritt wurden die Behandlungsverfahren nach den Bezeichnungen und Beschreibungen der StudienautorInnen den im Vorfeld festgelegten Kategorien zugeordnet. Diese Kategorien wurden auf der Basis von Metaanalysen (Bradley et al., 2005; Powers et al., 2010; Seidler & Wagner, 2006) und eines Übersichtsartikels (Cloitre, 2009) aufgrund der Annahme zusammengestellt, dass genügend Primärstudien für diese Kategorien vorhanden sind. Das Kriterium, um eine eigene Metaanalyse

für eine Kategorie von Verfahren durchführen zu können, war, dass mindestens fünf Studien pro Kategorie vorhanden sein mussten (Cuijpers et al., 2008). Eine eigene Kategorie wurde auch für alle als bona fide qualifizierten non-direktiven Verfahren gebildet. Nach dieser ersten Kategorisierung wurden in einem zweiten Schritt Kategorien zusammengefasst, um durch die größere Anzahl von inkludierten Studien auch eine größere Teststärke zu erzielen. Wenn nicht genügend Studien vorhanden waren, um eine eigene Kategorie zu bilden, wurden diese Studien für die Kategorienvergleiche nicht berücksichtigt. Eine alternative Kategorisierung in traumafokussierte und nicht-traumafokussierte Verfahren wurde ebenfalls vorgenommen. Nun folgen einige Details zur Kategorisierung.

5.3.1 EMDR

Da EMDR als Traumatherapiemethode erst 1987 entwickelt und in Buchform 1995 herausgegeben wurde (Shapiro, 1995), bestand großes Interesse, die Wirksamkeit des Verfahrens zu überprüfen und mit anderen traditionellen Psychotherapien zu vergleichen. EMDR ist ein in acht Phasen strukturiertes Verfahren, dessen zentrales Element spezifische Augenbewegungen sind, die während des Erinnerns an eine traumatische Szene durchgeführt werden. Trotz zahlreicher Komponentenstudien ist es noch immer nicht geklärt, ob die Augenbewegungen ein tatsächlich notwendiger Bestandteil des Verfahrens sind. Ein weiterer wesentlicher Bestandteil ist die Umwandlung negativer Glaubenssätze in positive Selbstrepräsentationen (Chemtob, Tolin, van der Kolk, & Pitman, 2000). Von den sieben Primärstudien, die 2006 in einer Metaanalyse zum Vergleich von EMDR mit TF-CBT inkludiert waren (Seidler & Wagner, 2006), sind sechs auch Teil dieser Arbeit. Der annähernd idente Studienkorpus ermöglichte einen Vergleich der Ergebnisse beider Analysen.

5.3.2 Expositionsverfahren und Prolonged Exposure

Die Begegnung mit der Erinnerung (Münker-Kramer, 2006) ist das zentrale Element der Expositionsverfahren, wobei diese in unterschiedlichen Formen stattfinden kann (in vivo, imaginativ, virtuell). *Prolonged Exposure* (PE) ist ein spezifisches Programm, das in dem von Foa, Rothbaum, Riggs und Murdock (1991) entwickelten Manual beschrieben wurde und eine Kombination von imaginativer und in vivo Exposition darstellt. Expositionsverfahren als eigene Kategorie wurden in der Metaanalyse

von Bradley et al. (2005) analysiert, wobei PE zu einem späteren Zeitpunkt in einer eigenen Metaanalyse auf seine Wirksamkeit untersucht wurde (Powers et al., 2010). Dieser Methodik folgend wurden in dieser Arbeit ebenfalls alle anderen Expositionsverfahren getrennt von PE untersucht, die beiden Kategorien aber auch in einer eigenen Metaanalyse zusammengefasst. In der Zuordnung der Primärstudien zur Kategorie PE bestehen allerdings Unterschiede zu Powers et al.:

- Ironson et al. (2002) verglichen in ihrer Studie EMDR mit PE. Um die Definition von Powers et al. zu erfüllen, sollten im Rahmen von PE sowohl in vivo als auch imaginative Expositionselemente durchgeführt werden. Ihrer Einschätzung nach fehlte die in vivo Exposition, sodass sie diese Studie aus ihrer Metaanalyse ausschlossen. Laut Beschreibung der StudienautorInnen wurden jedoch in vivo Expositionsaufgaben als Hausaufgaben durchgeführt.
- In drei Studien erfüllten Therapieverfahren die Definition von Powers et al. und wurden daher in ihre Metaanalyse inkludiert, die StudienautorInnen beschrieben selbst die Verfahren nicht als PE. Taylor et al. (2003), die ihr Verfahren als Expositionstherapie bezeichneten, führten zum Unterschied zur Therapiebeschreibung von Foa et al. (1991) die in vivo Expositionselemente nicht nur als Hausübungen durch, sondern widmeten die Hälfte aller Sitzungen der von den TherapeutInnen begleiteten in vivo Exposition. Die Kombination von Expositionsverfahren und kognitivem Umstrukturieren (McDonagh et al., 2005; Power et al., 2002) wurde in der vorliegenden Arbeit auch nicht als PE, sondern als CBT kodiert, da die kognitiven Elemente in der Therapiebeschreibung von Foa et al. nicht erwähnt wurden. McDonagh et al. (2005) bezeichneten ihre Therapie als CBT und als erweiterte Version der PE.

5.3.3 CBT

Laut den Richtlinien über effektive Behandlungsverfahren für PTSD (Rothbaum, Meadows, Resick, & Foy, 2000) werden acht verschiedene Therapieformen und Kombinationen davon der CBT zugeordnet. Unter diesen befinden sich Expositionsverfahren, das Stressimpfungstraining, die Kognitive Therapie und die Kognitive Prozesstherapie. Weiters werden Kombinationen dieser Therapieformen erwähnt, wie beispielsweise die Verbindung von Exposition und Kognitiver Therapie. Während Bradley et al. (2005)

in ihrer Metaanalyse diese Kombination als eine Kategorie verwendeten, und um eine zweite Kategorie ergänzten, die alle CBT-Verfahren enthielt, in denen keine Exposition durchgeführt wurde, fassten andere ForscherInnen kognitiv-verhaltenstherapeutische Verfahren unter der Kategorie TFEBT zusammen, sofern diese ihren Fokus auf die Bearbeitung des Traumas richteten (Bisson et al., 2007; Seidler & Wagner, 2006).

In dieser Arbeit wurden jene Therapien als CBT kodiert, die nicht in erster Linie Expositionsverfahren waren oder den Kategorien PE beziehungsweise EMDR zugeordnet werden konnten. Das bedeutet, dass aber in manchen Verfahren durchaus Expositionselemente angewandt wurden. Die Therapieformen dieser Kategorie sind detailliert unter 7.2.3 angeführt. Im Wesentlichen sind es folgende: Stressimpfungstraining, Kognitive Therapie und Kognitive Prozesstherapie, Kombinationen von Exposition und Kognitiver Therapie, die Narrative Expositionstherapie, eine traumafokussierte verhaltenstherapeutische Gruppentherapie, das traumaspezifische Therapieverfahren *Seeking Safety* (Hien et al., 2004) und das Therapieverfahren zur Behandlung posttraumatischer Albträume *Imagery Rehearsal* (Cook et al., 2010). Das Stressimpfungstraining, das zumeist unter dem Begriff Stressmanagement subsumiert wird (Bisson et al., 2007), und zwei weitere verhaltenstherapeutische Verfahren dieser Kategorie sind jedoch nicht als traumafokussiert zu bezeichnen (*Sleep & Nightmare Management*, Cook et al., 2010; *Relapse Prevention*, Hien et al., 2004).

5.3.4 Gegenwartszentrierte Therapie (PCT, Present-Centered Therapy)

Alle non-direktiven Verfahren, die als bona fide qualifiziert wurden, wurden dieser Kategorie zugeordnet. Gemeinsame Elemente fast aller dieser Therapieformen waren ein problemlösender Ansatz, die Fokussierung auf aktuelle Schwierigkeiten in Beziehungen oder persönlichen Lebensbereichen, Psychoedukation und der Hinweis auf eine klientenzentrierte Gesprächshaltung, wie sie beispielsweise Rogers (1973) empfohlen hat. Die Behandlungsverfahren dieser Kategorie können ausnahmslos als nicht-traumafokussiert bezeichnet werden.

5.3.5 Traumafokussierte versus nicht-traumafokussierte Verfahren

Die Interventionen der Kategorie PCT wurden um die nicht-traumafokussierten CBT-Verfahren ergänzt, sodass in dieser Kategorie ein direkter Vergleich zwischen diesen beiden Verfahrensgruppen durchgeführt werden konnte.

5.4 Moderatoren

In den Metaanalysen zur Wirksamkeit von PTSD-Behandlungsverfahren wurden bisher folgende Moderatoren untersucht:

- **Zum Trauma:** Art des Traumas (Bradley et al., 2005; Powers et al., 2010; Van Etten & Taylor, 1999), Zeit seit dem Trauma (Powers et al., 2010), Ausgangssymptomatik (Karatzias et al., 2007).
- **Zum Studiendesign:** Publikationsjahr (Bradley et al., 2005; Powers et al., 2010), Anzahl der Exklusionskriterien, Completer-Quote (Bradley et al., 2005), Studienqualität (Bisson et al., 2007; Powers et al., 2010), Randomisierung, Symptomkategorien, Selbstbeurteilungsfragebogen oder Interview, Art der Kontrollgruppe (Taylor & Harvey, 2009).
- **Zum Behandlungsverfahren:** Anzahl der Sitzungen gesamt (Powers et al., 2010; Taylor & Harvey, 2009), Individual- oder Gruppenverfahren, Anzahl der Sitzungen pro Woche, Dauer der Sitzungen, Strukturiertheit des Verfahrens, Klinik oder allgemeiner Kontext, Involviertheit des Therapeuten, Erfahrung der TherapeutInnen, Therapierichtung (Taylor & Harvey, 2009).
- **Zu den PatientInnen:** Geschlecht (Bisson et al., 2007).

Es ließen sich Hinweise finden, dass größere Effekte vorliegen bei: Individualtherapien, Erhebung per Interview, StudentInnen (meistens HochschulabsolventInnen oder DoktorandInnen) als TherapeutInnen, intensiveren und kürzeren Therapien (zweimal pro Woche für fünf bis acht Wochen) und strukturierten statt unstrukturierten Therapieverfahren (Taylor & Harvey, 2009). Größere Effekte ließen sich auch in jüngeren Studien finden, je mehr Exklusionskriterien eine Studie hatte, und je weniger PatientInnen eine Behandlung beendeten. Bezüglich der Art des Traumas fanden sich die niedrigsten Effekte bei Krieg und die größten bei Gewalttaten (Bradley et al., 2005). Die Effektgrößen waren bei Frauen als Patientinnen höher und bei Vietnamveteranen niedriger (Bisson et al., 2007). Auf der Suche nach signifikanten Prädiktoren zur Vorhersage von Ergebnissen dreier PTSD-Messinstrumente wurden bessere Ergebnisse bei einer niedrigeren Ausgangssymptomatik der PTSD, einer geringeren Anzahl von Sitzungen und Frauen als PatientInnen erzielt (Karatzias et al., 2007).

Auf Basis dieser Ergebnisse wurden in der vorliegenden Arbeit für folgende Variablen Moderatorenanalysen durchgeführt (siehe auch 6.6):

- **Zum Trauma:** Zeit seit dem Trauma, Schweregrad der PTSD-Ausgangssymptomatik.
- **Zum Studiendesign:** Studienqualität (siehe 5.4.1), ITT- oder Completeranalyse, homogene oder heterogene Zielgruppe (PatientInnen mit unterschiedlichen Traumata oder mit einem spezifischen Trauma).
- **Zur Therapie:** Anzahl der Sitzungen, Anzahl der Therapiewochen.
- **Zu den PatientInnen:** Geschlecht, Frauenanteil.

5.4.1 Studienqualität

Die Studienqualität wurde nach zwei verschiedenen Verfahren bewertet. Zum einen wurden die sieben Standards gut kontrollierter Studien der *Practice Guidelines from the International Society for Traumatic Stress Studies* (Foa et al., 2000) als Basis für eine Qualitätsbeurteilung herangezogen:

1. Eindeutig definierte Zielsymptomatik
2. Reliable und valide Messverfahren
3. InterviewerInnen, die blind für die Therapiebedingung sind
4. Training der diagnostischen InterviewerInnen
5. Manualisierte, replizierbare und spezifische therapeutische Programme
6. Unverzerrte Zuteilung zu den Therapiebedingungen
7. Einhaltung des Therapieprotokolls

Da zu den Einschlusskriterien der vorliegenden Arbeit die Randomisierung sowie die Diagnose der PTSD gehörte, und für die primären Maße Messinstrumente verwendet wurden, die in den Practice Guidelines angegeben waren, wurden diese Standards durch andere Kriterien einer qualitativ hohen Methodik ergänzt (Cuijpers et al., 2008), sodass schließlich folgende sieben Kriterien bewertet wurden:

- Die Randomisierung wurde in einem adäquaten Verfahren von unabhängiger Stelle (z.B. versiegelte Umschläge, Computerprogramme) durchgeführt.
- Die Erhebungen wurden blind für die Therapiebedingung durchgeführt.

- Die Gründe der Personen, die aus den Therapien ausgestiegen sind, wurden beschrieben, oder es wurden Vergleiche mit den Completern durchgeführt.
- Die TherapeutInnen waren oder wurden im jeweiligen Behandlungsverfahren trainiert.
- Qualifizierte Maßnahmen zur Einhaltung des Therapieprotokolls wurden getroffen beziehungsweise wurde die Einhaltung überprüft.
- ITT-Analysen wurden durchgeführt und verwendbare Daten berichtet.
- Es lag ein Manual oder strukturiertes Prozedere als Therapieleitfaden vor.

Diese Qualitätsbeurteilung ermöglichte es, die Studienqualität als kontinuierliche Variable für eine Moderatorenanalyse einzusetzen, während das folgende Studienqualitätsrating als kategoriale Moderatorvariable verwendet wurde.

Das kurze, reliable und valide Verfahren zur Qualitätsbeurteilung von randomisierten, kontrollierten Studien nach Jadad et al. (1996) wird in der Medizin angewandt, und wurde auch in der Psychotherapieforschung bereits in abgewandelter Form verwendet (Powers et al., 2010; Tolin, 2010). Folgende drei Kriterien werden dabei beurteilt, und für jedes wird ein Punkt vergeben:

1. **Adäquate Randomisierung:** Durch das Zuteilungsverfahren sollte jeder Patient die gleiche Ausgangslage haben, einem der Behandlungsverfahren zugeordnet zu werden, und die MitarbeiterInnen sollten nicht die Möglichkeit haben, vorherzusagen, welchem Behandlungsverfahren als nächstes zugeteilt wird. Wenn ein Randomisierungsverfahren in diesem Sinne durchgeführt wurde, wird ein Zusatzpunkt vergeben, andernfalls wird der Punkt für Randomisierung wieder abgezogen.
2. **Verblindung:** In der Originalversion wird hier ein *Double blinding*-Vorgehen verlangt, in dem sowohl die PatientInnen als auch die EvaluatorenInnen blind für das Verfahren sind. Da dies bei einem psychologischen Behandlungsverfahren nicht durchführbar ist, galt in der vorliegenden Arbeit nur die Verblindung der EvaluatorInnen als Kriterium.

3. **Dropouts:** Anzahl und Gründe jener Personen sollten genannt werden, die vorzeitig aus der Therapie ausgestiegen sind. Ergänzend zur Originalversion wurde dies auch als erfüllt betrachtet, wenn statistische Vergleiche zwischen Dropouts und Completern durchgeführt wurden.

5.5 Fehlende Werte

Die Basis für die Effektgrößenberechnung der kontinuierlichen Variablen waren Mittelwerte und Standardabweichungen. In zwei Studien wurde allerdings ein Teil der Daten nur in graphischer Form dargestellt (Rothbaum et al., 2005; Taylor et al., 2003), sodass diese so genau als möglich geschätzt wurden. Für die Metaanalyse von Seidler und Wagner (2006) wurden diese fehlenden Werte von den StudienautorInnen zur Verfügung gestellt, aber leider nicht in Form von Mittelwerten und Standardabweichungen veröffentlicht. Den publizierten Effektgrößen dieser Metaanalyse ist zu entnehmen, dass sich die Werte etwas von den in der vorliegenden Arbeit berechneten Werten unterscheiden. Im Vergleich zu den Effektgrößen von Benish et al. (2008), die auf die Einbeziehung dieser Maße ganz verzichtet haben, sind die Unterschiede jedoch geringer (siehe Tabelle 4).

Tabelle 4

Vergleich von Effektgrößen der primären Maße in Bezug auf den Umgang mit fehlenden Werten

Studie	Schätzwert in dieser Arbeit auf Basis der Graphiken ^a	Seidler und Wagner (2006) nach persönlicher Mitteilung ^a	Benish et al. (2008) ohne diese Maße ^{bc}
Taylor et al. (2003)	-0.59	-0.73	-0.11
Rothbaum et al. (2005)	-0.34	-0.42	-0.31

Anmerkungen. Die Ergebnisse dreier Messinstrumente wurden nur in graphischer Form berichtet. Je nach Vorgehen unterscheiden sich die Effektgrößen in den drei Metaanalysen.

^aHedges *g*. ^bCohens *d*. ^cDie Effektgrößen basieren auf anderen (sekundären) Ergebnismaßen.

Wenn Standardabweichungen nicht berichtet wurden, dann wurden die Effektgrößen mit dem *Practical Meta-Analysis Effect Size Calculator* (Wilson, 2010) anhand von Mittelwerten und dem Standardfehler berechnet. Wenn nur Effektgrößen in den Primärstudien berichtet wurden, dann wurden diese übernommen.

6 Statistische Methodik

Als Basisliteratur für die statistische Methodik, nach der die Effektgrößenberechnung und die metaanalytischen Verfahren erfolgt sind, wurde *Introduction to Meta-Analysis* (Borenstein et al., 2009) benutzt.

6.1 Effektgrößenberechnung

Um eine Metaanalyse rechnen zu können, wird im ersten Schritt für jede inkludierte Primärstudie eine Effektgröße und die dazugehörige Varianz berechnet. Die Effektgröße steht für die Beziehung zweier Variablen zueinander. Wird, wie in dieser Arbeit, die Wirksamkeit zweier Behandlungsverfahren miteinander verglichen, dann gibt die Effektgröße den Unterschied zwischen den beiden Verfahren in der jeweiligen Studie an. Jede Effektgröße wird mit einem p -Wert und einem Konfidenzintervall angegeben. Wenn das 95% Konfidenzintervall Null enthält, dann liegt der p -Wert über 0.05 und die Nullhypothese, dass es keinen Unterschied zwischen den Therapien gibt, kann nicht zurückgewiesen werden. Je enger das Konfidenzintervall ist, desto präziser wurde die Effektgröße in einer Studie geschätzt. Für die kontinuierlichen Variablen wurde der Standardisierte Mittelwertsunterschied (SMD, Standardized Mean Difference) als Effektgröße verwendet, für binäre Daten das Relative Risiko (RR). Jede Effektgröße steht für den direkten Vergleich zweier Bona Fide-Psychotherapien. In einer Studie (Brom et al., 1989) wurden drei Therapieverfahren miteinander verglichen, sodass sich für diese Studie drei Vergleiche ergaben ($A - B$, $B - C$, $A - C$). Die daraus resultierende Abhängigkeit wurde im Rahmen der Sensitivitätsanalysen berücksichtigt. Die Variablen, für die Metaanalysen gerechnet wurden, sind unter 5.2 beschrieben.

6.1.1 Standardisierter Mittelwertsunterschied: Cohens d , Hedges g

Da in den Primärstudien unterschiedliche Messinstrumente zur Erhebung der PTSD und relevanter psychischer Konstrukte eingesetzt wurden, war es notwendig, eine Effektgröße zu verwenden, die über die Studien hinweg vergleichbar ist. Nach Hedges und Olkin (1985) erfüllt die SMD diese Bedingung, da sie ein Maß dafür ist, wie weit zwei Verteilungen überlappen. Cohens d gibt den Abstand zweier Verteilungen in Einheiten ihrer gemeinsamen Standardabweichung an. Ein d von 0.5 bedeutet, dass der Mittelwert der ersten Therapiegruppe um eine halbe Standard-

abweichung größer ist als der Mittelwert der zweiten Therapiegruppe. Nach Cohen (1988) werden Effektgrößen von 0.8 als große Effekte, von 0.5 als mittlere Effekte und von 0.2 als kleine Effekte bezeichnet. Für die Berechnung von d wird der Mittelwert der einen Gruppe vom Mittelwert der anderen Gruppe subtrahiert und das Ergebnis durch die gepoolte Standardabweichung beider Gruppen dividiert. Diese Berechnungen wurden für die vorliegende Arbeit mit dem *Practical Meta-Analysis Effect Size Calculator* (Wilson, 2010) durchgeführt.

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_{pooled}}$$

In kleinen Stichproben wird d allerdings systematisch überschätzt und sollte daher korrigiert werden. Die daraus resultierende Effektgröße wird als Hedges g bezeichnet (Hedges, 1981). Um d zu g zu konvertieren, wurde folgender Korrekturfaktor J von Borenstein et al. (2009) vorgeschlagen, der eine Annäherung an die Formel von Hedges (1981) darstellt:

$$J = 1 - \frac{3}{4df - 1}$$

Die Freiheitsgrade df sind in dieser Formel $n_1 + n_2 - 2$ für zwei unabhängige Gruppen. Da der Korrekturfaktor J immer kleiner ist als 1.0, ist g immer kleiner als d und korrigiert damit die systematische Überschätzung von d . Laut Borenstein et al. (2009) ist es anzuraten, diesen Korrekturfaktor zu benutzen, auch wenn es außer in sehr kleinen Studien wenig praktischen Unterschied macht. Lipsey und Wilson (2001) gaben als kleine Studie eine Stichprobengröße von 20 an. Auch in den bereits erwähnten Metaanalysen zu PTSD-Behandlungsverfahren wurde zumeist Hedges g verwendet (Powers et al., 2010; Seidler & Wagner, 2006; Sherman, 1998; Taylor & Harvey, 2009) ebenso wie in den Practice Guidelines from the International Society for Traumatic Stress Studies (Foa et al., 2000). Benish et al. (2008) verwendeten diesen Korrekturfaktor nicht.

Die Varianz von g wurde nach folgender Formel berechnet (Viechtbauer, 2007):

$$\hat{\sigma}_g^2 = \frac{1}{\tilde{n}} + \left(1 - \frac{(m-2)}{m[c(m)]^2} \right) g^2$$

wobei $\tilde{n} = n_1 n_2 / (n_1 + n_2)$, $m = n_1 + n_2 - 2$ und $c(m) = 1 - \frac{3}{4m-1}$ sind.

Da die Mehrzahl der Primärstudien mehrere Maße zur Symptomerhebung verwendeten, wurde zuerst pro Messinstrument eine Effektgröße berechnet. Diese Effektgrößen wurden dann gemittelt, sodass pro Studie und Variable nur ein Effekt in die Analyse einging. Wenn in einer Studie für ein Messinstrument sowohl der Gesamtwert als auch die Werte der Subskalen angeführt wurden, dann wurde die Effektgröße nur für den Gesamtwert berechnet und die Subskalenwerte außer Acht gelassen. Die Abhängigkeit der Messinstrumente bei der Errechnung der gemittelten Effektgröße wurde wie bei Benish et al. (2008) mit einer Korrelation von 0.50 bei der Mittelung der Varianzen berücksichtigt, was genauere Schätzwerte zur Folge hat. Die Formel dazu (Borenstein et al., 2009, S. 228):

$$\text{var}\left(\frac{1}{m} \sum_{i=1}^m \right) = \left(\frac{1}{m}\right)^2 \left(\sum_{i=1}^m V_{i_1} + \sum_{i \neq j} (r_{ij} \sqrt{V_i} \sqrt{V_j}) \right)$$

Für jene Studien, in denen nur die Werte der Subskalen eines Messinstruments angeführt wurden, wurden zuerst in einem eigenen Rechenschritt die Effektgrößen und Varianzen für die Subskalen berechnet und diese dann gemittelt, sodass als Resultat eine Effektgröße und die dazugehörige Varianz für das Messinstrument in den zweiten Rechenschritt (die Mittelung der verschiedenen Ergebnismaße) einging. Für die Berechnung der Varianzen für die Subskalen wurden die publizierten Interkorrelationen der Subskalen dieser Messinstrumente verwendet (z.B. 0.52 – 0.82 für die Subskalen der Impact of Event Scale – Revidierte Form; Creamer, Bell, & Failla, 2003, S. 1491).

6.1.2 Relatives Risiko

Das Relative Risiko ist eine Effektgröße, die auf dichotomen Daten beruht und das Verhältnis der Ereignisraten zweier Vergleichsgruppen quantifiziert. Sowohl für den Vergleich der Dropout-Raten der beiden Therapiebedingungen als auch für den Vergleich der klinischen Signifikanz wurde das Relative Risiko als Effektgröße verwendet. Die Berechnung beruht auf einer 2 x 2 Tabelle (siehe Tabelle 5).

Tabelle 5

2x2 Tabelle zur Berechnung des Relativen Risikos

	Ereignis ^a	Nicht-Ereignis ^b	N
Behandlungsverfahren I	A	B	n_1
Behandlungsverfahren II	C	D	n_2

Anmerkungen. Für die Dropout-Analyse und die Berechnung der Klinischen Signifikanz wurde jeweils definiert, was Ereignis und was Nicht-Ereignis ist.

^aAnzahl der Dropouts bzw. Anzahl der TeilnehmerInnen ohne PTSD-Diagnose nach der Behandlung. ^bAnzahl der Completer bzw. Anzahl der TeilnehmerInnen mit PTSD-Diagnose nach der Behandlung.

Die Formel für die Berechnung des RR ist

$$RR = \frac{A/n_1}{C/n_2}$$

Da die Berechnungen des RR auf einer logarithmischen Skala basieren, ist

$$\text{Log}RR = \ln(RR)$$

mit der Varianz

$$V_{\text{Log}RR} = \frac{1}{A} - \frac{1}{n_1} + \frac{1}{C} - \frac{1}{n_2}$$

Diese Berechnungen wurden im Rahmen der Metaanalyse mit dem Programm R und dem Softwarepaket Metafor (Viechtbauer, 2010) durchgeführt. Das vom Programm ausgegebene $\text{Log}RR$ wurde in einem zweiten Schritt in die originale Skala zurückverwandelt. Ein RR von 1 bedeutet, dass es keinen Unterschied zwischen den Bona Fide-Psychotherapien gibt.

6.2 Metaanalyse und statistisches Modell

Sind alle Effektgrößen der Primärstudien bestimmt, wird in weiterer Folge die Metaanalyse als Synthese der einzelnen Effektgrößen gerechnet. Das Ergebnis ist der gewichtete Mittelwert der einzelnen Effektgrößen, Summeneffekt oder mittlerer Effekt genannt. Gewichtet wird jede einzelne Effektgröße durch ihre Präzision, ausgedrückt durch den Gewichtungsfaktor $1/\text{Varianz}$. Diese Gewichtung wurde vom Pro-

gramm R im Rahmen der Metaanalyse durchgeführt (Viechtbauer, 2010). Der mittlere Effekt wird in der graphischen Darstellung durch die Form eines Diamanten veranschaulicht. Die Lage des Diamanten repräsentiert die Effektgröße, die Weite des Diamanten die Präzision des Wertes, das Konfidenzintervall. Die statistische Signifikanz des mittleren Effekts wird auf der Basis des p -Wertes festgestellt, der wie bei den Primärstudien dann signifikant ist, wenn das Konfidenzintervall nicht Null enthält. Wenn alle inkludierten Therapieverfahren gleich wirksam sind, sollten die Summeneffekte niedrig sein und die p -Werte über .05 liegen, da die Nullhypothese dann nicht zurückgewiesen werden kann.

Die Interpretation des mittleren Effekts erfolgt je nach statistischem Modell, auf dem die Metaanalyse basiert. Dem *Fixed-Effect-Modell* liegt die Annahme zu Grunde, dass sich alle Studien in der Analyse die gleiche wahre Effektgröße teilen, und alle Unterschiede in den beobachteten Effekten nur aufgrund des Stichprobenfehlers zustande kommen. Somit ist im Fixed-Effect-Modell der mittlere Effekt der Schätzwert dieser wahren Effektgröße. Allerdings unterscheiden sich viele in einer Metaanalyse zusammengefassten Studien sowohl methodisch (z.B. hinsichtlich der Studienqualität) als auch in Bezug auf die Zusammensetzung des Stichprobe oder die Konzeption der Behandlungsverfahren, sodass die Verwendung des Fixed-Effect-Modells nicht angezeigt erscheint.

Unter dem *Random-Effects-Modell* hingegen wird angenommen, dass sich die wahre Effektgröße von Studie zu Studie unterscheidet. Die Effekte der in der Metaanalyse verwendeten Studien werden als zufällige Stichprobe von Effektgrößen angesehen, die um einen Mittelwert verteilt sind. Der Summeneffekt ist somit der Schätzwert für diesen Mittelwert. Die Nullhypothese ist, dass dieser Mittelwert Null (SMD) beziehungsweise 1.00 (RR) ist. Um eine Aussage über die Verteilung der wahren Effektgrößen machen zu können, werden im Random-Effects-Modell auch Vorhersageintervalle angegeben, in deren Bereich der wahre Effekt einer neuen Studie in 95% der Fälle liegt. Zum Unterschied dazu bedeutet das Konfidenzintervall, dass in 95% der Fälle der Mittelwert der wahren Effekte innerhalb dieses Intervalls liegt.

Borenstein et al. (2009) wiesen ausdrücklich darauf hin, dass es zwar eine gängige Praxis ist, eine Metaanalyse mit dem Fixed-Effect-Modell zu beginnen und nur im Falle eines statistisch signifikanten Tests auf Heterogenität das Random-

Effects-Modell zu verwenden, diese Vorgehensweise aber abzulehnen ist, da die Entscheidung, welches Modell einer Metaanalyse zugrunde gelegt wird, aufgrund inhaltlicher Überlegungen erfolgen sollte. Für die meisten Metaanalysen ist die Annahme, dass alle Studien in einer Analyse strukturell ident sind, nicht haltbar. Außerdem könnte nach dem Fixed-Effect-Modell keine Generalisierung auf eine breitere Population vorgenommen werden. Viechtbauer (2010) betonte sogar, dass dies das ausschlaggebende Argument für die Wahl des statistischen Modells sein sollte, nämlich ob man über die inkludierten Studien hinaus generalisieren will oder nicht. Allerdings ist zu berücksichtigen, dass bei sehr geringer Primärstudienanzahl die Präzision der Variation zwischen den Studien (T^2) im Random-Effects-Modell eingeschränkt ist. Wenn T^2 Null ergibt, dann entspricht die Random-Effects-Analyse einer Fixed-Effect-Analyse. Dieser Argumentation folgend wurden alle Metaanalysen mit dem Random-Effects-Modell durchgeführt.

6.3 Heterogenität

Abgesehen von der Größe und der Präzision des Summeneffekts ist eine weitere wichtige Information, wie die Effektgrößen der Primärstudien um den gemeinsamen Mittelwert verteilt sind. Diese Variation in den wahren Effektgrößen nennt man Heterogenität. Wenn sich alle Einzelstudien die gleiche wahre Effektgröße teilen, dann ist die Heterogenität gleich Null, wobei aber berücksichtigt werden muss, dass in den beobachteten Werten dennoch eine gewisse Variation aufgrund des Zufallsfehlers vorhanden ist. Somit ist es von Interesse, ob das Ausmaß der Heterogenität über den Zufallsfehler hinausgeht und eine Variation zwischen den Studien repräsentiert. Dies wird mit einem statistischen Signifikanztest geprüft. Für den Test auf Heterogenität werden die Q Statistik und der dazugehörige p -Wert herangezogen.

Die Q Statistik ist eine χ^2 -verteilte Prüfgröße, die nicht als absoluter Wert der wahren Variation interpretiert werden darf, sondern deren signifikanter p -Wert bedeutet, dass eine Variation in den wahren Effekten vorhanden ist. Ein nicht signifikanter p -Wert kann sowohl bedeuten, dass keine Variation vorhanden ist, oder aber auch, dass die Teststärke aufgrund einer geringen Anzahl von Studien und/oder einem großen Stichprobenfehler bei kleinen Studien zu gering ist. Über den statistischen Test hinaus geben folgende Werte Aufschluss über das Ausmaß der Heterogenität:

- T^2 , der Schätzwert des Parameters τ^2 (Varianz der wahren Effektgrößen), ist in der gleichen Maßeinheit wie die Effektgröße angegeben (z.B. SMD oder RR) und kann als absoluter Betrag der Variation angesehen werden.
- T ist der Schätzwert des Parameters τ (Standardabweichung der wahren Effektgrößen) und wird in Kombination mit dem Summeneffekt verwendet, um analog zu Mittelwert und Standardabweichung bei Einzelstudien die Verteilung der Effektgrößen um den mittleren Effekt zu beschreiben. Wenn beispielsweise der Summeneffekt 0.5 beträgt und $T = 0.10$, dann liegen die meisten Effektgrößen im Bereich zwischen 0.30 und 0.70 ($M \pm 2SD$).
- I^2 drückt den Anteil der wahren Heterogenität im Verhältnis zur gesamten Varianz aus. I^2 ist eine deskriptive Statistik und wird in Prozenten angegeben. Als Richtwerte werden angenommen (Higgins, Thompson, Deeks, & Altman, 2003): 0% = keine wahrgenommene Heterogenität, 25% = geringe Heterogenität, 50% = mittlere Heterogenität, 75% = hohe Heterogenität. Das bedeutet, dass bei einem I^2 von 50% die Hälfte der Variabilität zwischen den Effektgrößen der inkludierten Studien durch wahre Heterogenität und nicht durch den Stichprobenfehler verursacht wird (Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006). Wenn alle inkludierten Psychotherapien gleich wirksam sind, sollten die I^2 Werte im geringen Bereich liegen. Wenn die I^2 Werte jedoch hoch sind, dann können Subgruppenanalysen oder Metaregressionen dazu beitragen, die Varianz aufzuklären.

Bei der Interpretation von I^2 und dem Q -Test ist zu berücksichtigen, dass die Teststärke bei nur wenigen inkludierten Studien ($k < 20$) gering ist (Huedo-Medina et al., 2006). Daher können im Bedarfsfall Konfidenzintervalle für I^2 über die Genauigkeit der Schätzung der wahren Heterogenität Auskunft geben. In der Simulationsstudie von Huedo-Medina et al. wurde auch festgestellt, dass, wenn g als Effektgröße für den Vergleich von einem aktiven Verfahren und einer Kontrollbedingung benutzt wird, sowohl I^2 als auch der Q -Test problematisch zu interpretieren sind, da der Typ I Fehler dann sehr groß ist. Da in dieser Arbeit aber Direktvergleiche von aktiven Therapieverfahren analysiert wurden, trifft diese Einschränkung nicht zu. In Bezug auf die geringe Teststärke beider Indizes sind signifikante Ergebnisse des Q -Tests umso bedeutsamer.

6.4 Metaanalytische Verfahren

Um die Frage nach den Unterschieden zwischen Bona Fide-Psychotherapien auch in Hinblick auf die metaanalytische Methodik beantworten zu können, wurden drei verschiedene Verfahren verwendet, um die Effektgrößen auszuwerten. Zwei Verfahren basieren auf der Metaanalyse von Wampold et al. (1997), während es sich beim dritten Verfahren um ein metaanalytisches Standardverfahren handelt.

6.4.1 Random Signs-Verfahren

In der ersten von Wampold et al. (1997) veröffentlichten Metaanalyse zur vergleichenden Effektivität von Bona Fide-Psychotherapien wurde ein neues metaanalytisches Verfahren vorgestellt, das es ermöglicht, auf jegliche Kategorisierung von Behandlungsverfahren zu verzichten. Im ersten Schritt wird dabei den Effektgrößen der inkludierten Studien per Zufallsgenerator ein positives oder negatives Vorzeichen zugeordnet, da es für das Verfahren keine Rolle spielt, welche der beiden Therapien den größeren Effekt hat. Der resultierende Summeneffekt ist für die Analyse nicht von Bedeutung, da aufgrund der Vorzeichenverteilung ohnedies ein mittlerer Effekt um Null zu erwarten ist. Von Bedeutung ist allerdings der Test auf Heterogenität, der nach Aussage von Wampold et al. die Frage nach der vergleichenden Wirksamkeit der Behandlungsverfahren aufklären kann.

However, if the true effect size is zero, then the effect sizes should be distributed around zero, as predicted by sampling theory for effect sizes (i.e., many effects near zero and few effects in the tails of the distribution). If the true effect sizes is not zero, then there should be a disproportionate number of effects in the tails of this distribution. (S. 208)

Die Heterogenitätsanalysen in den Metaanalysen von Wampold und KollegInnen (Benish et al., 2008; Imel et al., 2008; Miller et al., 2008; Wampold et al., 1997) resultierten alle in nicht signifikanten Ergebnissen, sodass die Nullhypothese, dass die Therapieverfahren gleich wirksam sind, nicht zurückgewiesen werden konnte.

Es ist aber vorstellbar, dass die Zufallsverteilung der Vorzeichen einen Einfluss auf den Summeneffekt und die Verteilungsform hatte. Um festzustellen, ob das Ergebnis des Tests auf Heterogenität stabil ist, wurde in der vorliegenden Arbeit die Zu-

fallsverteilung der Vorzeichen mit dem *Excel Random Number Generator* 30-mal durchgeführt und mit jedem Ergebnis eine Metaanalyse gerechnet. Stimmt die Annahme, dass tatsächlich keine Heterogenität in den Studien vorhanden ist, sollte bei einem Fehlerniveau von 5% lediglich in etwa 2 (exakt: 1.5) von 30 Durchgängen ein signifikantes Ergebnis resultieren (dies lässt sich mithilfe der Binomialverteilung, einer Wahrscheinlichkeitsverteilung, mittels der die Anzahl der Erfolge in dieser Serie von gleichartigen und unabhängigen Versuchen bestimmbar ist, folgern; vgl. Bortz, 1993).

Es wurde die Benish et al. (2008) Metaanalyse mit genau jenen 15 Studien für die primären Maße nach dieser Vorgehensweise repliziert. Darüber hinaus wurde das Verfahren mit dem aktualisierten Studienkorpus für die primären und die sekundären Maße durchgeführt, da die größere Studienanzahl einen adäquateren Test auf Heterogenität ermöglichte (Higgins & Thompson, 2002).

6.4.2 Upper Bound-Verfahren

Die zweite Strategie, die in Wampold et al. (1997) verwendet wurde, beruht darauf, eine Metaanalyse mit den absoluten Werten der Effektgrößen zu rechnen, das heißt ohne Vorzeichen. Der Summeneffekt ist dann notwendigerweise größer als Null und stellt für Wampold et al. einen sogenannten *Upper Bound*, eine oberste Grenze des Summeneffekts dar. Ein Beispiel aus Messer und Wampold (2002): Angenommen die Behandlungsverfahren A und B sind gleich effektiv, aber aufgrund des Stichprobenfehlers ergibt sich in einer Studie ein Behandlungseffekt von $d = 0.20$ und in einer anderen Studie ein Effekt von $d = -0.20$. Die wahre Effektgröße wäre Null, aber der mittlere Effekt, mit den absoluten Werten gerechnet, wäre $d = 0.20$, also eine Überschätzung des wahren Effekts. „Absolute values do not take account of the sampling error that typically produces these small but opposite effects, which one would expect to occur even if the true difference were zero“ (S. 18).

Die Obergrenze von $d = 0.19$, die Wampold et al. (1997) errechnet haben, wird von manchen Kritikern als das wahre Maß für die relative Effektivität von Psychotherapie gesehen (Howard et al., 1997). Ein Upper Bound wurde in dieser Arbeit für die primären und sekundären Maße errechnet.

6.4.3 Kategorienvergleiche

Eine umfassende Metaanalyse zum Thema relative Effektivität wurde 2008 zum Störungsbild der Depression veröffentlicht (Cuijpers et al., 2008). In dieser Studie wurden Kategorien in einem ExpertInnenverfahren erarbeitet und ausschließlich Primärstudien inkludiert, die direkte Vergleiche von mindestens zwei Behandlungsverfahren für Depressionen berichteten. Insgesamt wurden 53 Studien und sieben Kategorien identifiziert, und für jede Kategorie wurde eine eigene Metaanalyse durchgeführt, in der die Therapieverfahren dieser Kategorie mit anderen Verfahren verglichen wurden. Zumindest fünf Studien waren notwendig, um eine Kategorie zu bilden. Dieses Vorgehen wurde für die vorliegende Arbeit übernommen, die Kategorienbildung erfolgte allerdings wie unter 5.3 beschrieben. Im Unterschied zu Cuijpers et al. (2008), die nur ein störungsspezifisches Maß und das *RR* für Dropouts analysierten, wurden in dieser Arbeit Analysen für die primären und sekundären Maße als auch für die Variablen Klinische Signifikanz und Dropouts durchgeführt. Diese Analysen wurden wenn möglich für mehrere Erhebungszeitpunkte gerechnet.

6.5 Sensitivitätsanalysen

Die Ergebnisse der einzelnen Metaanalysen wurden mittels dreier Arten von Sensitivitätsanalysen auf Robustheit überprüft.

6.5.1 Abhängigkeiten

Aufgrund von drei Vergleichen in einer Studie war eine Abhängigkeit in den Daten vorhanden (Brom et al., 1989). Deshalb wurden alle relevanten Metaanalysen noch einmal mit nur einem Vergleich, nämlich dem für die Nullhypothese herausforderndsten (= größeren) Effekt gerechnet. Weiters wurden in drei Studien zwei Verfahren einer Kategorie miteinander verglichen (Bryant et al., 2008; Cook et al., 2010; Hien et al., 2004), sodass auch für diese Kategorien die Metaanalysen noch einmal ohne die betreffenden Studien durchgeführt wurden.

6.5.2 Publication Bias

Da eine Metaanalyse eine Synthese aller inkludierten Einzelstudien ist, würde eine systematisch verzerrte Auswahl von Studien zu einem systematischen Fehler im Summeneffekt führen. Eine solche Verzerrung könnte das Ergebnis einer einseitigen

Suchstrategie sein, allerdings wurde auch bereits mehrfach untersucht, ob nicht eine systematische Verzerrung schon bei der Veröffentlichung von Forschungsergebnissen auftritt (Dickersin, Chan, Chalmersx, Sacks, & Smith Jr., 1987). Tatsächlich wurde nachgewiesen, dass eher Studien veröffentlicht werden, die statistisch signifikante Ergebnisse und damit meist größere Effekte berichten, und dass diese schneller publiziert werden als Studien, die keine signifikanten Ergebnisse aufweisen. Diesem unter dem Namen *Publication Bias* bekannten Problem kann zweifach begegnet werden: Einerseits mit einer Suchstrategie, mit der auch nach unveröffentlichten Untersuchungen aus dem Bereich der Grauen Literatur wie Dissertationen oder Forschungsberichte gesucht wird, und andererseits mit Verfahren, die die Ergebnisse einer Metaanalyse auf einen Publication Bias hin untersuchen und die Effekte möglichst unverzerrt schätzen. Von Borenstein et al. (2009) wurden dazu drei verschiedene Möglichkeiten beschrieben:

Erstens kann man anhand eines Trichterdiagramms feststellen, ob es in der Verteilung der Effektgrößen überhaupt einen Hinweis auf einen Publication Bias gibt. Da es Grundannahmen darüber gibt, welche Studien nicht publiziert werden, könnte man diese anhand der Graphik überprüfen. Es sind vor allem Studien mit einer geringen Stichprobengröße, die in Gefahr laufen, nicht veröffentlicht zu werden, da bei diesen Studien nur sehr große Effekte signifikant werden. Große Studien hingegen werden so gut wie immer publiziert und Studien von mittlerer Größe werden auch meistens publiziert, da bei dieser Stichprobengröße auch mittlere Effekte signifikant werden. Diesen Annahmen zufolge wäre ein Publication Bias erkennbar, wenn die Verteilung im Trichterdiagramm an der Spitze des Trichters symmetrisch wäre, in der Mitte wenige Studien fehlen würden, und am Boden des Trichters die Symmetrie deutlich verletzt wäre.

Die zweite Möglichkeit ist, Orwins *Fail-safe N* zu berechnen, um festzustellen, wie viele fehlende Studien den Summeneffekt auf einen substantiell anderen Wert bringen würden. Somit kann mit diesem Verfahren festgestellt werden, ob der Summeneffekt nur ein Artefakt des Publication Bias ist. Während die Interpretation des Trichterdiagramms durchaus subjektiv ist, und Orwins *Fail-safe N* nur die Aussage ermöglicht, ob der Publication Bias in die Überlegungen miteinbezogen werden muss, erlaubt das dritte Verfahren eine differenzierte Betrachtung der Problematik.

Mittels des *Trim-and-fill* Verfahrens (Duval & Tweedie, 2000) kann berechnet werden, wie viele Studien wahrscheinlich fehlen, als auch ein Schätzwert der unverzerrten Effektgröße unter Berücksichtigung dieser fehlenden Studien bestimmt werden. Dieses Verfahren kann somit die Robustheit der Ergebnisse in Hinblick auf einen Publication Bias hin überprüfen und wurde in der vorliegenden Arbeit für die Kategorienvergleiche verwendet.

6.5.3 Einflussreiche Fälle

Zusätzlich wurde ein Verfahren auf die Kategorienvergleiche angewandt, das es ermöglicht, Ausreißer beziehungsweise einzelne einflussreiche Fälle zu identifizieren (*Leave-One-Out*-Verfahren; Viechtbauer, 2010). Als einflussreich werden jene Fälle bezeichnet, wenn es hinsichtlich des mittleren Effekts oder der Heterogenität zu bedeutsamen Änderungen kommt, würde man diesen Fall aus der Analyse entfernen.

6.6 Moderatoranalysen

Analog zu einer Varianzanalyse und zur Multiplen Regressionsanalyse bei Primärstudien können in Metaanalysen die mittleren Effekte von Subgruppen miteinander verglichen werden und der Einfluss von Moderatorvariablen auf die Effektgröße berechnet werden (Borenstein et al., 2009). Moderatoranalysen werden eingesetzt, um zumindest einen Teil der Heterogenität in den Daten aufzuklären (Viechtbauer, 2010).

Mittels Subgruppenanalyse ist beispielsweise feststellbar, ob der Unterschied zwischen zwei Behandlungsverfahren in Studien von geringer methodischer Qualität größer ist als in qualitativ hochwertigen Studien. Für die Subgruppenanalyse werden kategoriale Moderatoren eingesetzt, und es können zwei oder mehr Kategorien gebildet werden. Als Ergebnis erhält man den mittleren Effekt für jede Subgruppe sowie die dazugehörigen p -Werte und Konfidenzintervalle. Ein Ergebnis könnte daher sein, dass in einer Subgruppe ein signifikanter Unterschied zwischen den Verfahren vorhanden ist, während in einer oder mehreren anderen Subgruppen die p -Werte nicht signifikant werden. Dies entspricht einer differenzierenden Vorgehensweise, denn mögliche Unterschiede zwischen den Behandlungsverfahren könnten nicht erkannt werden, wenn alle Effekte in einer einzigen Analyse zusammengefasst werden.

In Übereinstimmung mit dem Vorgehen von Cuijpers et al. (2008) wurden in der vorliegenden Arbeit Subgruppenanalysen nur für jene Kategorien von Behandlungsverfahren durchgeführt, in denen mindestens 10 Effekte enthalten waren.

Folgende kategorialen Variablen wurden analysiert (siehe auch 5.4):

- Geschlecht (weiblich – männlich – gemischt),
- Zielgruppe der Studie (unterschiedliche Traumata – spezifisches Trauma),
- Analyse (ITT oder Completer),
- Studienqualitätsrating nach Jadad et al. (1996),
- Schweregrad der PTSD (schwere – extrem schwere Symptombelastung).

Um den Zusammenhang zwischen einer kontinuierlichen Moderatorvariable und dem mittleren Effekt zu bestimmen, wurden Metaregressionen gerechnet. Es wird ebenfalls empfohlen, dass pro Moderatorvariable 10 Studien in der Analyse enthalten sein sollten (Borenstein et al., 2009). Als Ergebnis erhält man folgende Informationen: Hat die Moderatorvariable einen signifikanten Einfluss und reduziert sie die Heterogenität? Ist nach Berücksichtigung des Moderators noch immer signifikante Heterogenität vorhanden?

Kontinuierliche Moderatorvariablen in dieser Arbeit waren:

- Frauenanteil in Prozenten,
- Studienqualität nach 7 Kriterien,
- Zeit seit dem Trauma,
- Anzahl der Sitzungen,
- Anzahl der Therapiewochen.

Für die statistische Analyse wurde jeweils ein *Mixed-Effect*-Modell verwendet. Das bedeutet, dass die Metaanalyse für alle Effekte mit dem Random-Effects-Modell durchgeführt wurde und der mittlere Effekt innerhalb der Subgruppen beziehungsweise die Metaregression mit dem Fixed-Effect-Modell berechnet wurden.

6.7 Statistisches Programm

Das statistische Programm, das für die Durchführung der Metaanalysen, des Trim-and-fill Verfahrens und der Moderatorenanalysen verwendet wurde, ist das Programm R (Copyright © 2011 The R Foundation for Statistical Computing) in der Version 2.13.0 (2011-04-13). Es handelt sich um eine freie Software, die unter <http://www.R-project.org> heruntergeladen werden kann. Das Datenpaket, das für die Metaanalysen benötigt wird, heißt Metafor (Meta-Analysis Package for R) und wurde in der Version 1.6-0 benutzt (Viechtbauer, 2010).

7 Deskriptive Ergebnisse

7.1 Neue Studien

Zusätzlich zu den 15 Studien, die schon im Studienkorpus von Benish et al. (2008) enthalten waren, erfüllten sieben weitere Studien die Einschlusskriterien. Insgesamt wurden 115 Abstracts gesichtet und 35 Studien im Original überprüft. Eine Auflistung von Studien, die in die engere Wahl kamen, aber wesentliche Kriterien nicht erfüllten, findet sich im Anhang. Nach Ehlers et al. (2010) sollten diese in Metaanalysen aus Gründen der Nachvollziehbarkeit aufgelistet werden. Darüber hinaus könnte man auch feststellen, inwieweit sich Ergebnisse unter Berücksichtigung dieser Studien ändern, wenn auch die Effektgrößen angegeben würden. Dies war der Fall bei einer Studie, die zuerst in den Studienkorpus aufgenommen wurde, aber aufgrund fehlender Diagnostik der PTSD wieder ausgeschlossen werden musste (Classen et al., 2011). Welche Auswirkungen das Inkludieren dieser Studie auf die Ergebnisse hatte, wird unter 9.2.1 diskutiert.

Da in vier von den sieben neuen Studien jeweils ein Behandlungsverfahren den gegenwartszentrierten Therapien und damit der umstrittenen Kategorie der non-direktiven oder supportiven Verfahren (Ehlers et al., 2010; Wampold et al., 2010) zuzuordnen war, wurden diese Verfahren hinsichtlich der Bona Fide-Definition detailliert analysiert. In allen vier Therapien führten ausgebildete TherapeutInnen die Sitzungen durch, eine PatientInnen-TherapeutInnen-Beziehung konnte etabliert werden, und es bestand die Möglichkeit, die Behandlungsverfahren an die PatientInnen zu adaptieren. Somit konzentriert sich die folgende Analyse auf jene vier Kriterien, von denen zwei für die Bona Fide-Definition erfüllt sein mussten (Benish et al., 2008; Wampold et al., 2010): die Zitierung eines etablierten psychologischen Verfahrens; die Beschreibung des Behandlungsverfahrens auf Basis psychologischer Prinzipien; das Vorhandensein eines Manuals als Leitfaden für die Therapie und das Nennen aktiver Bestandteile des Behandlungsverfahrens mit Quellenangabe.

Neuner, Schauer, Klaschik, Karunakara und Elbert (2004) - *Supportive Counselling*: Es war leider nicht festzustellen, ob diese Studie für Benish et al. (2008) die Bona Fide-Definition nicht erfüllte oder ob sie nicht über diese Studie verfügten. Zwei Aussagen in der Beschreibung dieser Therapieform sind kritisch zu betrachten:

Die Kontrolle auf nicht-spezifische Behandlungseffekte als Hauptzweck des Verfahrens und der Hinweis, dass das Verfahren möglicherweise nicht repräsentativ für die Praxis war, da es vermieden wurde, über das Trauma zu sprechen. Für Wampold et al. (2010) sind Behandlungsverfahren unter diesen Bedingungen nicht therapeutisch intendiert. Die Therapiebeschreibung entsprach aber den geforderten zwei Kriterien. Einerseits wurde eine Zitierung eines psychologischen Beratungsansatzes gemacht, der als Grundlage des Verfahrens diene (A trauma counselling handbook by Amani Trust, 1997; zitiert nach Neuner et al., 2004, S. 583), andererseits wurden typische psychologische Bestandteile von gegenwartszentrierten Verfahren genannt: Ressourcenorientierung, Fokussierung auf interpersonelle Probleme, Entscheidungen und Zukunftsgestaltung und das Erlernen eines Problemlöseansatzes. Zusätzlich fand eine Sitzung Psychoedukation statt, und den TeilnehmerInnen wurde ein Rational für die Behandlung vermittelt.

Schnurr et al. (2007) - *Present-Centered Therapy*: Auch in dieser Studie wurde der Zweck des Verfahrens mit der Kontrolle auf nicht-spezifische Therapieeffekte angegeben, wobei in diesem Fall das Verfahren als realistisch eingeschätzt wurde. Die StudienautorInnen verwiesen mit Quellenangabe (Rosen et al., 2004) darauf, dass diese Therapie typischerweise in der Praxis von PsychologInnen und PsychotherapeutInnen verwendet wird, und dass es sich um ein glaubwürdiges Therapieverfahren handelt. Dieses Argument wurde von Wampold et al. (2010) kritisch hinterfragt. Ihrer Meinung nach ähnelte das Verfahren, das in der Studie durchgeführt wurde, nur rudimentär jenen gegenwartszentrierten Verfahrensweisen, die in der zitierten Praxis tatsächlich angewandt werden. Bewertet man strikt die Bona Fide-Kriterien, so war ein Manual als Therapieleitfaden vorhanden, und es gab eine kurze Beschreibung: Behandlungsrational, Psychoedukation, Fokussierung auf gegenwärtige Lebensprobleme, non-direktive Gesprächsführung. Diese Beschreibung ist durchaus kritisch zu sehen, und man könnte das Verfahren sicher auch als nicht-bona fide einstufen, was Wampold et al. (2010) auch getan haben.

In der vorliegenden Arbeit wurde dennoch die Entscheidung getroffen, das Verfahren als bona fide zu qualifizieren, weil in einer Studie aus dem Studienkorpus von Benish et al. (2008) ein ähnliches Verfahren als Gruppentherapie verwendet wurde (Schnurr et al., 2003). Es erschien widersprüchlich, von der gleichen Forsche-

rInnengruppe ein Verfahren einmal ein- und einmal auszuschließen, nur weil es sich in dem einen Fall um ein Individualverfahren und in dem anderen Fall um ein Gruppenverfahren handelte.

Abgesehen vom Format ähnelten sich die Beschreibungen in beiden Studien. Die von Benish et al. als bona fide inkludierte gegenwartszentrierte Gruppentherapie (Schnurr et al., 2003) wurde als nicht-spezifisches Verfahren zur Kontrolle der allgemeinen Faktoren beschrieben, wo es vermieden wurde, Traumabezüge herzustellen und kognitive Umstrukturierungen oder andere Komponenten der TFCBT durchzuführen. Das Verfahren wurde auch explizit als Kontrollbedingung bezeichnet (Schnurr, Friedman, Lavori, & Hsieh, 2001). Die Inhalte der Gruppentherapie waren Psychoedukation, Fokussierung auf interpersonelle Probleme, Problemlösung und Coping. Die Beschreibung war vielleicht um eine Nuance problemlösungsorientierter als in der Studie von Schnurr et al. (2007), aber hinsichtlich der konkreten Kriterien der Bona Fide-Definition war kein Unterschied erkennbar. Auch die angegebene Zusatzliteratur (Schnurr et al., 2001), lieferte nur eine zusätzliche Zitierung eines gruppentherapeutischen Modells. Das sollte allerdings nicht das Kriterium sein, da sonst fälschlicherweise der Schluss gezogen werden könnte, dass Gruppentherapieverfahren eher bona fide sind als Einzelverfahren. Die Schnurr et al. (2007) Studie ist aufgrund ihrer Stichprobengröße von über 140 Patientinnen pro Therapiebedingung von großer Bedeutung.

Cottraux et al. (2008) – Rogerian Supportive Therapy: Dieses Verfahren erfüllte die Kriterien eindeutig. Es war ein Manual vorhanden, das psychotherapeutische Verfahren nach Rogers wurde zitiert, und es wurden 10 Prinzipien aus der klientenzentrierten Psychotherapie als Bestandteile des Verfahrens genannt.

Ready, Gerardi, Backscheider, Mascaro und Rothbaum (2010) - Present-Centered Therapy: Als Kontrollbedingung entwickelt, erfüllte dieses Verfahren zwei der vier Kriterien für die Bona Fide-Definition. Einerseits wurden die gegenwartszentrierten Bestandteile beschrieben (Psychoedukation, Problemlösetechniken, Fokus auf das Hier und Jetzt, Entwicklung von Copingstrategien), und andererseits wurde auf die Haltung der klientenzentrierten Psychotherapie nach Rogers (1973) mit der entsprechenden Zitierung verwiesen.

7.2 Beschreibung aller inkludierten PTSD-Studien

Der Studienkorpus der vorliegenden Arbeit wird auf Studienebene, auf PatientInnenebene und auf Ebene der Behandlungsverfahren wie folgt beschrieben:

7.2.1 Studienstichprobe

Der Studienkorpus besteht aus insgesamt 22 Studien mit 24 Vergleichen, da in einer Studie drei Bona Fide-Psychotherapien miteinander verglichen wurden. 18 Studien wurden in englischsprachigen Ländern durchgeführt (11 in den USA, 3 in Großbritannien, 1 in Kanada, 3 in Australien), 3 Studien in Kontinentaleuropa und 1 Studie in Afrika. Alle Studien wurden in wissenschaftlichen Fachzeitschriften publiziert. In 12 Studien gab es keine Kontrollgruppen, während in 10 Studien zusätzlich zu den Behandlungsgruppen Kontrollbedingungen vorhanden waren.

In allen Studien wurde eine Erhebung zu einem Follow-up-Zeitpunkt durchgeführt. Das erste Follow-up fand im Zeitraum zwischen 1 und 15 Monaten nach Behandlungsende statt ($M = 4.67$ Monate). In neun Studien wurden auch Daten zu einem zweiten Follow-up erhoben. Dieses lag zwischen 6 und 24 Monaten nach Behandlungsende ($M = 10$ Monate).

Die Bewertung der Studienqualität wurde auf zweifache Weise durchgeführt (siehe 5.4.1). Einmal wurden sieben methodische Qualitätskriterien bewertet, wobei hoch qualitative Studien sechs oder sieben Kriterien erfüllten, einmal wurden nach einem Verfahren zur Qualitätsbeurteilung von randomisierten, kontrollierten Studien (Jadad et al., 1996) Punkte für drei Kriterien vergeben: die adäquate Randomisierung, die Verblindung und die Beschreibung der Dropouts. Nach diesem Verfahren waren insgesamt 4 Punkte für qualitativ hochwertige Studien zu erreichen.

In Tabelle 6 findet sich eine Übersicht, inwieweit die einzelnen Studien die sieben Qualitätskriterien erfüllten. Sechs Studien sind demnach als methodisch hoch qualitativ zu bezeichnen. Weiters findet sich in der Tabelle auch die Punkteanzahl, die die Studien nach dem Verfahren nach Jadad et al. (1996) erzielten. Sieben Studien erreichten hier die höchste Punkteanzahl. Bewertet wurden nur eindeutig erfüllte Kriterien. Die anderen Studien erfüllten die Kriterien nicht oder machten keine eindeutigen Angaben dazu.

Tabelle 6

Erfüllte Qualitätskriterien und Qualitätsbeurteilung der Primärstudien nach Jadad et al. (1996)

Qualitätsbeurteilung nach Jadad et al. (1996)								Pkte	
Studien	RAN	BLIN	DROP	THER	TREU	ITT	MAN	Anzahl	
Paunovic & Öst (2001)			x					1	1
Brom et al. (1989)			x	x				2	1
Deville & Spence (1999)			x	x	x			3	1
Ironson et al. (2002)			x		x		x	3	1
Lee et al. (2002)				x	x		x	3	0
Rothbaum et al. (2005)		x			x		x	3	1
Ready et al. (2010)		x	x				x	3	2
Marks et al. (1998)		x	x		x		x	4	2
Tarrier et al. (1999)	x	x			x		x	4	3
Hien et al. (2004)				x	x	x	x	4	0
Foa et al. (1991)		x	x	x	x		x	5	2
Foa et al. (1999)		x	x	x	x		x	5	2
Resick et al. (2002)			x	x	x	x	x	5	1
Taylor et al. (2003)		x	x	x	x		x	5	2
Neuner et al. (2004)	x	x	x	x	x			5	4
Cottraux et al. (2008)	x	x	x	x			x	5	4
Power et al. (2002)	x	x	x	x	x		x	6	4
Schnurr et al. (2003)	x	x	x		x	x	x	6	4
McDonagh et al. (2005)		x	x	x	x	x	x	6	2
Schnurr et al. (2007)	x	x	x	x	x	x	x	7	4
Bryant et al. (2008)	x	x	x	x	x	x	x	7	4
Cook et al. (2010)	x	x	x	x	x	x	x	7	4
Anzahl der Studien	8	15	18	15	18	7	18	-	-
Anteil der Studien in %	36.4	68.2	81.8	68.2	81.8	31.8	81.8	-	-

Anmerkungen. RAN = Randomisierung mittels adäquatem Verfahren; BLIN = Verblindung der Erhebungen; DROP = Beschreibung oder Vergleich der Dropouts; THER = Training der TherapeutInnen; TREU = Einhaltung des Therapieprotokolls; ITT = Verwendbare ITT-Analysen; MAN = Manual oder strukturiertes Prozedere; Anzahl = Anzahl der erfüllten Qualitätskriterien; Pkte = Punkte auf Basis der Qualitätsbeurteilung nach Jadad et al. (1996). Drei Kriterien wurden bewertet, und es konnten maximal vier Punkte erreicht werden.

7.2.2 PatientInnenstichprobe

Insgesamt 1694 PatientInnen mit einer PTSD nahmen an den Bona Fide-Psychotherapien teil. Im Schnitt waren das 77 PatientInnen pro Studie und rund 71 pro Vergleich. Folgende Charakteristika kennzeichneten die Stichprobe:

- **Frauenanteil:** 7 Studien hatten ausschließlich Frauen als Teilnehmerinnen (664 Personen), 3 Studien ausschließlich Männer (460 Personen) und 12 Studien Patienten beiderlei Geschlechts (570 Personen, Frauenanteil $M = 55.98\%$, Median = 60.12%).
- **Alter:** Der Altersschnitt in den einzelnen Studien reichte von 21.45 bis 59.42 Jahren mit einem Mittelwert von 38.6 Jahren (Median = 37.93).
- **Traumata:** In 10 Studien wurden Erwachsene mit unterschiedlichen Traumata (z.B. Verkehrsunfälle, kriminelle Handlungen, der Verlust von Angehörigen durch Gewalttaten oder schwere Erkrankungen) behandelt (519 Personen), während in 12 Studien die Wirksamkeit von Bona Fide-Psychotherapien für spezifische Zielgruppen untersucht wurde (1175 Personen):
 - in 5 Studien Opfer von Vergewaltigung und sexuellen Übergriffen in Kindheit oder Erwachsenenalter (305 Personen),
 - in 4 Studien Kriegsveteranen (744 Personen),
 - in 2 Studien Flüchtlinge (51 Personen),
 - in 1 Studie Patientinnen mit komorbidem Substanzmissbrauch (75 Personen).
- **Zeit seit dem Trauma:** In 14 Studien wurde explizit angegeben, vor wie vielen Jahren das traumatisierende Ereignis stattgefunden hat. Der kürzeste Zeitraum waren 1.24 Jahre, der längste 22.90 Jahre ($M = 7.14$, Median = 7.65).
- **Schweregrad:** Als schwer traumatisiert konnten die PatientInnen von 11 Studien (902 Personen) eingeordnet werden (8 Studien CAPS 60-79, 3 Studien IES > 43), 5 Studien (556 Personen) hatten als PatientInnenstichprobe extrem schwer traumatisierte Personen (CAPS ab 80). Die anderen Studien machten über den Schweregrad keine Angabe oder verwendeten Messinstrumente, die keine Schweregradbestimmung zuließen (PDS, PSS-I, PSS-SR, PCL).

- **Komorbiditäten:** Diese wurden nur in wenigen Studien angegeben:
 - in 5 Studien komorbide Angsterkrankungen mit einem PatientInnenanteil von 31% – 52.40% ($M = 41.20$),
 - in 8 Studien komorbide Depressionen mit einem PatientInnenanteil von 19.19% – 63.71% ($M = 48.22\%$) und
 - in 4 Studien komorbide substanzinduzierte Störungen mit einem PatientInnenanteil von 2.10% – 14.35% ($M = 7.01\%$). In 1 Studie war Substanzmissbrauch Einschlusskriterium, in 10 Studien war Alkohol- oder Substanzabhängigkeit ein Ausschlusskriterium.
- **Psychopharmakagebrauch:** In 11 Studien wurde der Prozentanteil jener PatientInnen angegeben, die Psychopharmaka einnahmen (19 – 94%, $M = 51.68$, Median = 45).
- **Rekrutierung:**
 - In 9 Studien wurden die PatientInnen im klinischen Umfeld rekrutiert (Überweisungen durch Ambulatorien, Ärzte oder Spitler, stationre PatientInnen; 1008 Personen),
 - In 1 Studie wurden die PatientInnen durch ffentliche Aufrufe mittels Inseraten oder Flugblttern gewonnen (60 Personen),
 - 6 Studien nutzten beide Rekrutierungswege (232 Personen),
 - in 1 Studie wurden PatientInnen von einer frheren Erhebung kontaktiert (31 Personen),
 - und in 5 Studien gab es dazu keine Angaben (363 Personen).

7.2.3 Inkludierte Behandlungsverfahren

In den 22 Primrstudien wurden insgesamt 45 Bona Fide-Psychotherapien durchgefhrt, die, wie in Tabelle 7 ersichtlich ist, folgenden Kategorien zugeordnet wurden: Expositionsverfahren, PE (Prolonged Exposure), CBT (Kognitive Verhaltenstherapie), EMDR (Eye Movement Desensitization and Reprocessing), PCT (Gegenwartszentrierte Therapie) und der Vergleich von traumafokussierten mit nicht-traumafokussierten Behandlungsverfahren. Fr diesen Vergleich wird in weiterer Folge die Abkrzung TF vs. NTF verwendet. Zustzlich wurden Expositionsverfahren und PE in einer Kategorie nochmals zusammengefasst.

Die Psychodynamische Kurzzeittherapie wurde als einziges Behandlungsverfahren keiner Kategorie zugeordnet. In früheren Metaanalysen war es meistens in die Gruppe der *anderen Verfahren* inkludiert worden (Bisson et al., 2007; Bradley et al., 2005), passte in der vorliegenden Arbeit aber weder in die Kategorie der gegenwartszentrierten Verfahren noch in die Kategorie der nicht-traumafokussierten Verfahren. Inhaltlich hatte die kurze Beschreibung, dass mit dieser Therapie innerpsychische Konflikte gelöst werden sollen, die durch die traumatische Erfahrung verursacht wurden (Brom et al., 1989), nicht viel mit den gegenwartszentrierten Verfahren gemein, und es ließ sich auch nicht ausschließen, dass das Trauma selbst Gegenstand der Therapie war. Das Verfahren „... is explicitly directed at the discontinuation of the present disorders and is not aimed at bringing about personality changes“ (Brom et al., S. 608). Diese Argumentation wurde auch von Wampold et al. (2010) ins Feld geführt, die die Einteilung in traumafokussierte und nicht-traumafokussierte Therapieverfahren aufgrund mangelnder objektiver Grundlage kritisierten. Darüber hinaus war auch die Wirksamkeit dieser Therapie nicht ganz eindeutig gegeben (siehe 2.4.2).

Auch die Zuordnung der Hypnotherapie (Brom et al., 1989) zu den Expositionsverfahren entspricht nicht früheren Einordnungen in die Gruppe der anderen Verfahren (Bisson et al., 2007; Bradley et al., 2005). Aus der Therapiebeschreibung ging jedoch hervor, dass das wesentliche Element eine Form von Exposition war:

The emphasis of the hypnotherapists in our study was on behavioral therapy. The goal was to bring the patient in contact with the reality of the traumatic event and to bring about a decrease in the conditioned responses triggered by the event. Hypnosis was used, because it allows flexibility in the way the client deals, both cognitively and emotionally, with the perception of and adjustment to the trauma. (Brom et al., 1989, S. 607)

Weitere Charakteristika der Therapieverfahren betrafen Format und Dauer. Nur vier Therapien (8.9%) wurden im Gruppenformat abgehalten, alle anderen waren Individualtherapien. Es fanden rund 12 Sitzungen ($M = 11.72$) zu 90 Minuten ($M = 85.50$) in durchschnittlich 11.07 Wochen statt. Das kürzeste Verfahren umfasste vier Sitzungen in zwei Wochen. Die längste Therapie dauerte 30 Wochen und umfasste 24 Sitzungen. Die Dropout-Rate betrug im Schnitt 21.48%.

Tabelle 7

Kategorienzuordnung der inkludierten PTSD-Behandlungsverfahren

		EX	PE	CBT	EMDR	PCT
Brom et al. (1989) <i>n</i> = 89	Trauma Desensitization	x				
	Psychodynamische Therapie ^a					
	Hypnotherapie ^b	x				
Foa et al. (1991)^c <i>n</i> = 31	PE		x			
	Stressimpfungstraining			x		
Marks et al. (1998) <i>n</i> = 42	PE		x			
	Cognitive Restructuring			x		
Foa et al. (1999)^c <i>n</i> = 51	PE		x			
	Stressimpfungstraining			x		
Deville & Spence (1999), <i>n</i> = 32	EMDR				x	
	Trauma Treatment Protocol			x		
Tarrier et al. (1999) <i>n</i> = 72	Imaginative Exposition	x				
	Kognitive Therapie			x		
Paunovic & Öst (2001), <i>n</i> = 20	Expositionstherapie	x				
	CBT			x		
Ironson et al. (2002) <i>n</i> = 22	EMDR				x	
	PE		x			
Lee et al. (2002) <i>n</i> = 26	Stressimpfungstraining + PE			x		
	EMDR				x	
Power et al. (2002) <i>n</i> = 76	EMDR				x	
	Exposure + Cog. Restructuring			x		
Resick et al. (2002) <i>n</i> = 124	Kognitive Prozesstherapie			x		
	PE		x			
Schnurr et al. (2003)^c <i>n</i> = 325	Trauma-focused Group Therapy			x		
	Present-centered Group Therapy					x
Taylor et al. (2003) <i>n</i> = 41	Expositionstherapie	x				
	EMDR				x	
Neuner et al. (2004)^c <i>n</i> = 31	Narrative Expositionstherapie			x		
	Supportive Counselling					x
Hien et al. (2004)^c <i>n</i> = 75	Seeking Safety			x		
	Relapse Prevention			x		
McDonagh et al. (2005)^c <i>n</i> = 51	CBT			x		
	Problem Solving Therapy					x

		EX	PE	CBT	EMDR	PCT
Rothbaum et al. (2005)	PE		x			
<i>n</i> = 48	EMDR				x	
Schnurr et al. (2007)^c	PE		x			
<i>n</i> = 284	Present-centered Therapy					x
Bryant et al. (2008)	Imaginative Exposition	x				
<i>n</i> = 59	In vivo Exposition	x				
Cottraux et al. (2008)^c	CBT			x		
<i>n</i> = 60	Rogierian Supportive Therapy					x
Cook et al. (2010)^c	Imagery Rehearsal			x		
<i>n</i> = 124	Sleep & Nightmare Management			x		
Ready et al. (2010)^c	Virtual Reality Exposure	x				
<i>n</i> = 11	Present-centered Therapy					x
Anzahl der Behandlungsverfahren pro Kategorie		8	7	17	6	6
PatientInnenanzahl pro Kategorie^d		221	300	705	123	376

Anmerkungen. EX = Expositionsverfahren; PE = Prolonged Exposure; CBT = Kognitive Verhaltenstherapie; EMDR = Eye Movement Desensitization and Reprocessing; PCT = Gegenwartszentrierte Therapie.

^aDiese Therapieform wurde keiner Kategorie zugeordnet. ^bDie Zuordnung basiert auf der Therapiebeschreibung der Studienautoren. ^cDiese Studien wurden in den Vergleich von traumafokussierten und nicht-traumafokussierten Verfahren aufgenommen. Gesamtanzahl: 10 Studien mit 1043 Personen. ^dbasiert auf den PatientInnenzahlen der zugeordneten Vergleiche.

7.3 Effektgrößen

Die Effektgrößen der primären und sekundären Maße der inkludierten Studien werden in Abbildung 2 und Abbildung 3 in Form eines *Forest Plots* dargestellt.

Einige Effektgrößen der primären PTSD-Maße unterscheiden sich von Benish et al. (2008) aus mehreren Gründen. Zum einen aufgrund der Variablenzusammensetzung (siehe 5.2), zum anderen wurde in zwei Studien unterschiedlich mit fehlenden Werten umgegangen (siehe 5.5). Weiters wurde in der vorliegenden Arbeit Hedges *g* als Effektgröße verwendet, während Benish et al. Cohens *d* benutzten (siehe 6.1.1).

Darüber hinaus wurde in zwei Studien eine andere Datenbasis zur Berechnung der Effektgrößen herangezogen (Foa et al., 1991; Ironson et al., 2002). In beiden Stu-

dien wurden Mittelwerte und Standardabweichungen zum Behandlungsende sowohl für die gesamte Completer-Stichprobe als auch für eine Teilstichprobe veröffentlicht, von der auch Follow-up-Daten zur Verfügung standen. In dieser Arbeit wurden die Werte der gesamten Completer-Stichprobe zur Effektgrößenberechnung herangezogen, bei Benish et al. die Werte der kleineren Stichprobe, von der auch Follow-up-Daten vorlagen. In Tabelle 8 wird der Unterschied in den Effektgrößen je nach zugrundeliegender Datenbasis veranschaulicht, wobei zur besseren Vergleichbarkeit nur Cohens *d* als Effektgröße angegeben wird.

Tabelle 8

Vergleich von Cohens d auf Basis unterschiedlicher Stichproben

Studie	Ergebnismaße	Completerstichprobe	Follow-up-Stichprobe
Foa et al. (1991)	PTSD Schweregrad	-0.56 ^a	-0.48 ^{be}
Ironson et al. (2002)	PTSD Maße	-0.65 ^c	-0.13 ^d

Anmerkungen. Completerstichprobe = Jene PatientInnen, die die Behandlung beendet haben. Follow-up-Stichprobe = Teilstichprobe derjenigen PatientInnen, von denen auch Follow-up-Daten zur Verfügung standen. PTSD = Posttraumatische Belastungsstörung.

^a $n_1 = 14, n_2 = 10$. ^b $n_1 = 9, n_2 = 9$. ^c $n_1 = 9, n_2 = 10$. ^d $n_1 = 6, n_2 = 6$. ^eDiese Effektgröße wurde gemäß der Variablenkonzeption noch mit anderen Effektgrößen gemittelt.

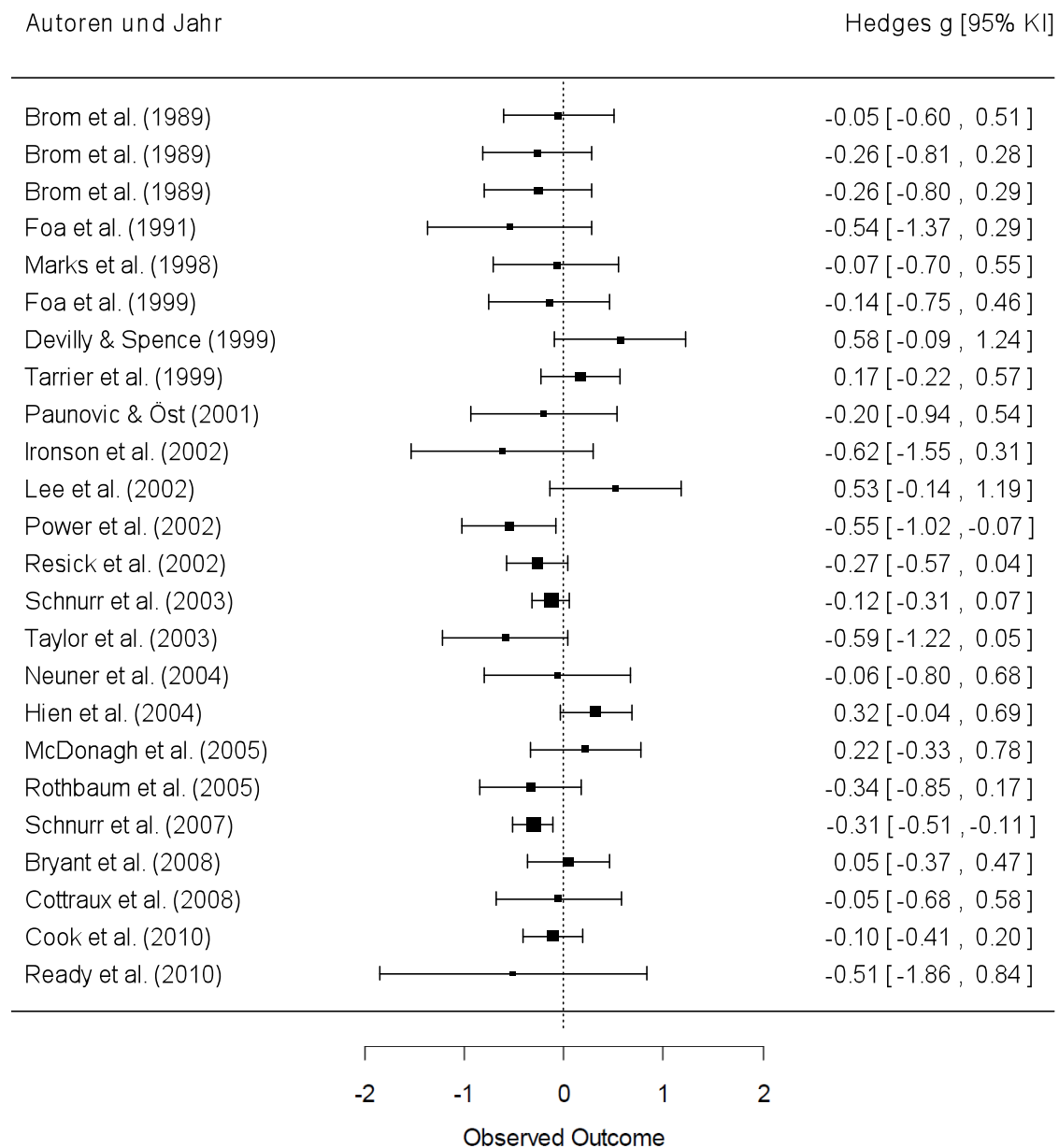


Abbildung 2. Effektgrößen der primären PTSD-Maße aller im Studienkorpus inkludierten Einzelstudien.

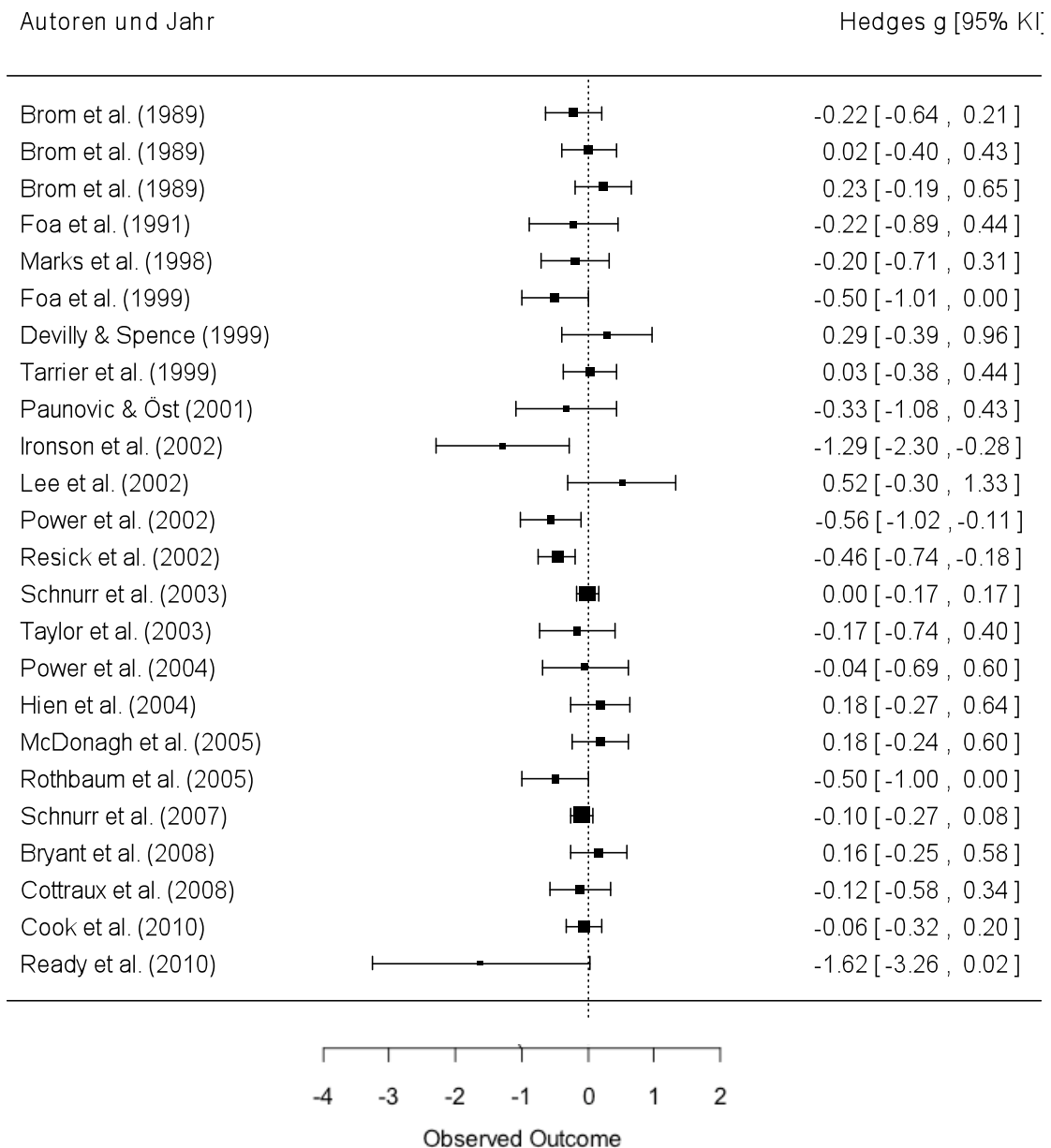


Abbildung 3. Effektgrößen der sekundären Maße aller im Studienkorpus inkludierten Einzelstudien.

8 Ergebnisse der Metaanalysen

8.1 Das Random Signs-Verfahren

Nach Wampold et al. (1997) geht es bei diesem Verfahren um den Test auf Heterogenität, der in Hinblick auf die Nullhypothese, dass alle Bona Fide-Psychotherapien gleich wirksam sind, Aufschluss geben kann. Da, wie unter 6.4.1 beschrieben, den Effektgrößen der inkludierten Studien per Zufallsgenerator negative oder positive Vorzeichen zugeordnet wurden, war ein mittlerer Effekt um Null zu erwarten, der aber in diesem Zusammenhang nicht aussagekräftig ist. Damit die Nullhypothese (homogene Verteilung der Effektgrößen um Null) nicht verworfen wird, darf das Ergebnis des Tests auf Heterogenität nicht signifikant ausfallen. Benish et al. (2008) erhielten in ihrer Metaanalyse zu den Behandlungsverfahren für PTSD als Ergebnis eine homogene Verteilung der Effektgrößen. Deshalb wurde im ersten Schritt das Verfahren für die primären PTSD-Maße jener 15 Studien wiederholt, die in dieser Metaanalyse inkludiert waren. Im zweiten Schritt wurde das Verfahren für alle im Studienkorpus enthaltenen 22 Studien sowohl für die primären PTSD-Maße als auch für die sekundären Maße angewandt. Die Zufallsverteilung der Vorzeichen wurde jeweils 30-mal durchgeführt. Somit wurden für jede Analyse 30 Durchgänge gerechnet, um feststellen zu können, ob die Ergebnisse, die aufgrund der zufälligen Vorzeichenzuordnung entstanden sind, stabil sind.

In Tabelle 9 sind die Ergebnisse der beiden ersten Durchgänge jeder Analyse exemplarisch angeführt. In der Analyse des Studienkorpus von Benish et al. (2008) ist der Test auf Heterogenität einmal signifikant und einmal nicht signifikant. Sowohl die p -Werte als auch die I^2 -Werte liegen in diesen beiden Durchgängen sehr eng zusammen. Von allen 30 Durchgängen gab es 10-mal ein nicht signifikantes und 20-mal ein signifikantes Ergebnis. Unter Annahme einer Binomialverteilung und einem Fehlerniveau von 5% für jedes einzelne Testergebnis ist diese Häufung signifikanter Ergebnisse selbst hochsignifikant (Binomialtest, $p < .001$). Der Erwartungswert signifikanter Ergebnisse lag unter diesen Annahmen lediglich bei 1.5. Das Heterogenitätsmaß I^2 lag im Schnitt aller 30 Durchgänge bei 38.24% und entsprach damit nicht dem Ergebnis von Null, das Benish et al. berechnet hatten.

In der Analyse der primären PTSD-Maße des gesamten Studienkorpus (22 Studien) ist der Test auf Heterogenität in den beiden ersten Durchgängen ebenfalls einmal signifikant und einmal nicht signifikant, allerdings ist das Ergebnis deutlicher, wenn alle 30 Durchgänge betrachtet werden. Insgesamt war der Q -Wert nur zweimal nicht signifikant. Das bedeutet, dass die Daten wiederum eindeutig nicht mit der Nullhypothese vereinbar sind ($p < .001$). I^2 lag im Schnitt aller 30 Durchgänge bei 40.30% und näherte sich dem Bereich der mittleren Heterogenität von 50% an.

Die Analyse der sekundären Maße bot ein einheitliches Bild. In allen 30 Durchgängen war der Q -Wert signifikant ($p < .001$). Der I^2 -Wert lag im Mittel bei 43.17%. Im Vergleich dazu berichteten Benish et al. (2008) für alle Maße ein nicht signifikantes Ergebnis des Heterogenitätstests und einen I^2 -Wert von 9.39%.

Tabelle 9

Ergebnisse des Random Signs-Verfahrens zur relativen Effektivität von Bona Fide-Therapien

Analyse	k	Mittelwertsvergleich			Heterogenität			
		ES	95% KI	<i>p</i>	<i>T</i>	<i>Q</i>	<i>p</i>	<i>I</i> ² (%)
Studienkorpus wie								
Benish et al. (2008)								
Primäre PTSD-Maße								
1. Durchgang	17	0.04	[-0.12, 0.20]	.660	0.198	26.85	.043	40.05
2. Durchgang	17	0.02	[-0.13, 0.18]	.757	0.190	25.72	.058	38.06
Studienkorpus gesamt								
Primäre PTSD-Maße								
1. Durchgang	24	-0.06	[-0.18, 0.06]	.349	0.163	34.96	.053	35.26
2. Durchgang	24	0.03	[-0.10, 0.15]	.661	0.183	39.34	.018	40.72
Sekundäre Maße								
1. Durchgang	24	0.02	[-0.09, 0.14]	.706	0.173	43.39	.006	43.26
2. Durchgang	24	0.003	[-0.11, 0.12]	.956	0.176	43.71	.006	44.28

Anmerkungen. k = Anzahl der Vergleiche; ES = Effektgröße Hedges g ; KI = Konfidenzintervall; T = Schätzwert der Standardabweichung der wahren Effektgrößen; $Q = \chi^2$ -verteilte Prüfgröße für den Test auf Heterogenität; I^2 = Heterogenitätsmaß in Prozenten. Primäre PTSD-Maße = Symptome der PTSD; Sekundäre Maße = Komorbide Symptomatik, traumabezogene Symptomatik und funktionsbezogene Maße. 1. bzw. 2. Durchgang = Die ersten beiden Ergebnisse aus 30 Durchgängen Zufallsverteilung von Vorzeichen im Rahmen des Random Signs-Verfahrens.

Zusammenfassend lässt sich folgendes festhalten:

Die Ergebnisse aller drei Analysen bestätigten nicht die Ergebnisse der Benish et al. (2008) Metaanalyse. Nach der Grundkonzeption des Random Signs-Verfahrens bedeuten die Ergebnisse der vorliegenden Arbeit, dass die Nullhypothese, dass alle Bona Fide-Psychotherapien gleich wirksam sind, zurückgewiesen werden muss, und dass das Dodo Bird Verdict nicht bestätigt werden kann (Benish et al., 2008; Wampold et al., 1997). Da sowohl der Studienkorpus von 2008 als auch der aktualisierte Studienkorpus analysiert wurden, lassen sich folgende Schlüsse ziehen:

- Nur eine einzige zufällige Vorzeichenzuteilung durchzuführen, liefert unter Umständen ein unzuverlässiges Ergebnis, das auf keinen Fall als Beleg für die Nullhypothese interpretiert werden sollte.
- Da sich die Effektgrößen zum Teil von Benish et al. unterscheiden (siehe 5.2, 5.5 und 7.3), ist es naheliegend, dass es für das Ergebnis dieses Verfahrens von wesentlicher Bedeutung ist, welche Ergebnismaße in die Variablen einfließen und von welcher Datenbasis die Effektgrößen berechnet werden.
- Die größere Anzahl von analysierten Effekten könnte auch aufgrund größerer Teststärke zu diesem Ergebnis beigetragen haben.

Ein Argument, das immer wieder in die Diskussion eingebracht wird, lautet, dass Behandlungsverfahren der inkludierten Studien nicht der Bona Fide-Definition genügen und dass, wenn man diese Studien aus dem Studienkorpus entfernt, die verbleibenden therapeutischen Verfahren gleich effektiv sind. Dieses Vorgehen wurde bei einer Metaanalyse zur Wirksamkeit von kognitiver Therapie bei Depressionen angewandt (Gloaguen et al., 1998; Wampold et al., 2002). Auch für die vorliegende Arbeit könnte dieses Argument ins Feld geführt werden, da von den sieben neuen Studien in vier Studien non-direktive Behandlungsverfahren verwendet wurden, die teilweise, wie unter 7.1 beschrieben, hinsichtlich der Bona Fide-Definition umstritten sind (Schnurr et al., 2007). Aber selbst unter Verwendung des Studienkorpus von Benish et al. (2008) war das Ergebnis des Tests auf Heterogenität eindeutig, sodass die Schlussfolgerungen bestehen bleiben.

8.2 Das Upper Bound-Verfahren

Zusätzlich zum Random Signs-Verfahren wurde auch ein mittlerer Effekt auf Basis der absoluten Werte der Effektgrößen berechnet, das heißt, die Effektgrößen gingen ohne Vorzeichen in die Analyse ein. Es sollte damit eine oberste Grenze des Summeneffekts bestimmt werden (Benish et al., 2008; Wampold et al., 1997), der nach diesem Vorgehen notwendigerweise größer als Null ist. Wie Tabelle 10 zu entnehmen ist, wurde dieses Verfahren sowohl für die primären PTSD-Maße als auch für die sekundären Maße zu allen drei Erhebungszeitpunkten durchgeführt.

Die Ergebnisse sind mit Benish et al. (2008) in etwa vergleichbar. Während in dieser Arbeit der mittlere Effekt für die primären PTSD-Maße zum Ende der Behandlung bei $g = 0.23$ und für die sekundären Maße bei $g = 0.20$ lag, berichteten Benish et al. einen mittleren Effekt von $d = 0.19$ für die PTSD-Symptome und $d = 0.16$ für alle Maße. Die geringen Unterschiede in den Werten sind in den unterschiedlichen Effektgrößen und nicht in den neuen Studien begründet, da der Upper Bound jener Studien, die in der Benish et al. Analyse enthalten waren, in der vorliegenden Analyse auch $g = 0.24$ betrug. In beiden Metaanalysen waren die Effektgrößen homogen verteilt.

In der Metaanalyse von Wampold et al. (1997) wurde der Upper Bound mit $d = 0.19$ berechnet und für eine Teilstichprobe von Studien mit DSM-Diagnosen ebenfalls mit $d = 0.23$. Dieser Wert um 0.23 scheint ein stabiles Ergebnis zu sein, da zur Überprüfung von Zeiteffekten eine kumulative Metaanalyse für die absoluten Effekte durchgeführt wurde, in der sich das Ergebnis ohne größere Veränderungen durch das Publikationsjahr bestätigte.

Auffällig in dieser Analyse sind die Maße zum zweiten Follow-up. Einerseits stiegen die Effektgrößen im Vergleich zum ersten Follow-up an, andererseits näherte sich der Test auf Heterogenität der Signifikanzgrenze, und die I^2 -Werte als Indikator für Heterogenität waren ebenfalls größer. Unter Umständen weist das auf viele Konfundierungen hin, die aufgrund des längeren Zeitraums vorhanden sind, da nicht kontrolliert werden kann, ob PatientInnen in dieser Zeit nicht auch andere Therapieangebote in Anspruch nehmen (Crits-Christoph, 1997). Ein weiterer Grund könnten die uneinheitlichen Erhebungszeitpunkte in den inkludierten Studien sein (Siev & Chambless, 2007).

Über alle Upper Bound-Analysen hinweg konnte man feststellen, dass die sekundären Maße immer etwas mehr Heterogenität aufwiesen als die primären PTSD-Maße. Dies könnte daran liegen, dass eine Vielzahl unterschiedlicher Konstrukte in den sekundären Maßen zusammengefasst wurde.

Da im Studienkorpus drei Vergleiche von einer Studie stammten (Brom et al., 1989), wurde die Robustheit der Ergebnisse überprüft (siehe 6.5). Tatsächlich änderte sich kaum etwas an den Ergebnissen (z.B. $g = 0.24$ statt $g = 0.23$).

Tabelle 10

Ergebnisse des Upper Bound-Verfahrens zur relativen Effektivität von Bona Fide-Therapien

Variable	k	Mittelwertsvergleich			Heterogenität			
		ES	95% KI	<i>p</i>	<i>T</i>	<i>Q</i>	<i>p</i>	<i>I</i> ² (%)
Behandlungsende								
Primäre PTSD-Maße	24	0.23	[0.15, 0.32]	<.0001	0	11.09	.982	0
Sekundäre Maße	24	0.20	[0.10, 0.29]	<.0001	0.092	25.63	.319	17.83
1. Follow-up								
Primäre PTSD-Maße	23	0.21	[0.12, 0.30]	<.0001	0	12.99	.933	0
Sekundäre Maße	23	0.19	[0.11, 0.28]	<.0001	0.052	19.17	.635	6
2. Follow-up								
Primäre PTSD-Maße	9	0.26	[0.11, 0.42]	.001	0.102	14.68	.066	18.48
Sekundäre Maße	9	0.24	[0.06, 0.42]	.009	0.173	15.00	.059	46.25

Anmerkungen. k = Anzahl der Vergleiche; ES = Effektgröße Hedges g ; KI = Konfidenzintervall; T = Schätzwert der Standardabweichung der wahren Effektgrößen; $Q = \chi^2$ -verteilte Prüfgröße für den Test auf Heterogenität; I^2 = Heterogenitätsmaß in Prozenten. Primäre PTSD-Maße = Symptome der PTSD; Sekundäre Maße = Komorbide Symptomatik, traumabezogene Symptomatik und funktionsbezogene Maße.

Die Detailanalysen zum Unterschied von Selbstbeurteilungs- und Interviewverfahren zeigten nur minimale Unterschiede in den Effektgrößen, während die Detailanalysen zu den sekundären Maßen interessante Hinweise auf Unterschiede in den Effektgrößen je nach Variable lieferten, wie in Tabelle 11 ersichtlich ist. Obwohl sich die Anzahl der Vergleiche, die in diese Analyse eingingen, unterschieden, und sich die Konfidenzintervalle überlappten, bestand dennoch zu zwei Erhebungszeitpunkten ein Unterschied zwischen traumabezogenen und funktionsbezogenen sekundären Maßen.

Die beiden Effektgrößen der funktionsbezogenen sekundären Maße waren knapp nicht signifikant und zeigten damit keinen Unterschied zwischen den Behandlungsverfahren an, obwohl aufgrund der Verwendung der absoluten Werte der Effektgrößen für die Analyse ein signifikanter Unterschied erwartet worden wäre. Dies kann als Hinweis darauf interpretiert werden, dass funktionsbezogene Maße getrennt von anderen sekundären Maßen analysiert werden sollten, da diese sonst mögliche Unterschiede in den sekundären Maßen maskieren könnten. Zum zweiten Follow-up zeigte sich auch in den Detailanalysen ein Anstieg an Heterogenität in den I^2 -Werten.

Tabelle 11

Upper Bound-Verfahren zur relativen Effektivität von Bona Fide-Therapien:

Detailanalyse der sekundären Maße

Variable	k	Mittelwertsvergleich			Heterogenität			
		ES	95% KI	<i>p</i>	<i>T</i>	<i>Q</i>	<i>p</i>	<i>I</i> ² (%)
Behandlungsende								
PTSD-bezogen	23	0.24	[0.15, 0.33]	<.0001	0.001	18.89	.652	0
Funktionsbezogen	11	0.07	[-0.03, 0.17]	.196	0	5.49	.856	0
1. Follow-up								
PTSD-bezogen	22	0.23	[0.14, 0.33]	<.0001	0	14.49	.848	0
Funktionsbezogen	10	0.10	[-0.00, 0.21]	.055	0	10.12	.341	0
2. Follow-up								
PTSD-bezogen	9	0.26	[0.08, 0.43]	.003	0.145	12.43	.133	35.84
Funktionsbezogen	7	0.27	[0.02, 0.51]	.032	0.194	9.52	.146	41.97

Anmerkungen. k = Anzahl der Vergleiche; ES = Effektgröße Hedges g ; KI = Konfidenzintervall; T = Schätzwert der Standardabweichung der wahren Effektgrößen; Q = χ^2 -verteilte Prüfgröße für den Test auf Heterogenität; I^2 = Heterogenitätsmaß in Prozenten. PTSD-bezogen = Depression, Angst, Substanz- und Alkoholkonsum, traumabezogene Symptome und Konstrukte wie Schuld, Ärger, Dissoziation und entsprechende Kognitionen; Funktionsbezogen = Maße wie gesundheitsbezogene Lebensqualität und soziales Funktionsniveau.

Zusammenfassung

- Das Upper Bound-Verfahren resultierte in stabilen Effekten um 0.20-0.23, wobei die Heterogenität (außer zum zweiten Follow-up) gegen Null ging. Tatsächlich ist der Upper Bound aber keine Obergrenze des wahren Effekts,

da es sich definitionsgemäß um einen mittleren Effekt handelt. Für die primären PTSD-Maße ($g = 0.23$) lagen Konfidenz- und Vorhersageintervall zwischen 0.15 und 0.32. Das bedeutet, dass in 95% der Fälle der Mittelwert der wahren Effekte und der wahre Effekt einer neuen Studie in diesem Bereich liegen. Eine Obergrenze wäre daher wohl eher mit 0.32 gegeben.

- Zu unterscheiden ist dies auch von der obersten Grenze eines möglichen Behandlungseffekts. Da der mittlere Effekt ein Mittelwert aller Effekte ist, sind naturgemäß einzelne Effekte wesentlich größer. Der größte absolute Effekt in der Stichprobe betrug 0.62, der niedrigste absolute Effekt 0.05. Hier stellt sich die Frage, ob der mittlere Effekt nicht vielmehr das Maß der Variabilität der Behandlungseffekte ist. Vergleicht man die Ergebnisse des Random Signs-Verfahrens, in dem mittlere Effekte um Null mit größtenteils signifikanten Heterogenitätstests resultierten, und die Ergebnisse des Upper Bound-Verfahrens, in dem keine Heterogenität und dafür Effekte um 0.23 vorhanden waren, so legt dies einen Zusammenhang nahe. Messer und Wampold (2002) erklärten als Grund dafür, warum der mittlere Effekt eine Überschätzung und nicht das wahre Maß des Unterschiedes ist, wie andere Autoren meinten (Howard et al., 1997), dass sich Schwankungen der Effektgrößen um Null aufgrund des Stichprobenfehlers wegen der absoluten Werte nicht ausmitteln würden (siehe 6.4.2). Die Problematik dieser Argumentation ist aber, dass sie voraussetzt, dass idente Vergleiche zwischen zwei Behandlungsverfahren entsprechend häufig und mit vergleichbarer Präzision in der Stichprobe zu finden sind, was tatsächlich nicht der Fall ist.
- Da in einer Stichprobe manche Vergleiche zwischen zwei Therapieverfahren häufiger vorkommen (wie der Vergleich von CBT und PCT in 4 von 22 Studien) und andere Vergleiche nur selten (beispielsweise der Vergleich von Hypnotherapie und Expositionsv erfahren in nur einer Studie) und manche dieser Vergleiche aufgrund der Stichprobengröße und der methodisch hoch qualitativen Durchführung der Studie ein größeres Gewicht haben, ergibt sich eine Unausgewogenheit der Stichprobe, die eine Verzerrung des Upper Bound bewirkt. Abgesehen davon, dass manche Vergleiche in der Stichprobe fehlen, weil sie nie durchgeführt wurden.

8.3 Kategorienvergleiche

Es wurden für folgende Einzelkategorien von Behandlungsverfahren eigene Metaanalysen durchgeführt: Expositionsverfahren, PE, CBT, PCT und EMDR. Zusätzliche Kategorien waren die Kombination von Expositionsverfahren und PE (EXPE) und der Vergleich von traumafokussierten mit nicht-traumafokussierten Therapien (TF vs. NTF). Um in weiterer Folge die Ergebnisse leichter vergleichbar zu machen, wurden die Vorzeichen aller mittleren Effekte so adaptiert, dass eine positive Effektgröße ein besseres Ergebnis für die analysierte Kategorie bedeutet.

8.3.1 Primäre und sekundäre Maße

Die Vergleiche der einzelnen Kategorien mit den anderen Bona Fide-Psychotherapien ergaben zwei signifikante Effekte. Die PCT waren signifikant weniger wirksam in den primären PTSD-Maßen als die Vergleichsverfahren, und zwar sowohl zum Behandlungsende als auch zum ersten Follow-up. Bei allen anderen Analysen konnte die Nullhypothese, dass Bona Fide-Psychotherapien keine Unterschiede in ihrer Wirksamkeit aufweisen, nicht zurückgewiesen werden.

Die Effektgröße für die Expositionsverfahren mit $g = 0.17$ ($p = .112$) scheint im Vergleich auch bedeutsam, selbst wenn sie knapp nicht signifikant geworden ist. Dass eine Differenzierung von Analysen Unterschiede eher hervortreten lässt (Siev et al., 2009), sieht man im Vergleich zur mit PE ($g = 0.01$, $p = .927$) erweiterten Kategorie EXPE, für die der mittlere Effekt nur mehr $g = 0.10$ ($p = .240$) betrug.

Wie Tabelle 12 zu entnehmen ist, zeigte der Test auf Heterogenität für zwei Kategorien sowohl bei den primären als auch bei den sekundären Maßen signifikante Ergebnisse. PE und EMDR wiesen ein beträchtliches Ausmaß an Heterogenität auf, die I^2 -Werte lagen mit 55.29%, 81.72%, 70.37% und 72.77% im mittleren bis hohen Bereich. Leider war die Studienanzahl in beiden Kategorien zu gering, um Moderatorenanalysen durchführen zu können. In der kombinierten Kategorie EXPE waren allerdings die sekundären Maße noch immer heterogen und lagen mit einem I^2 -Wert von 56.59% im mittleren Bereich. In dieser Kategorie wurden daher ebenso wie für die CBT Moderatorenanalysen zu den sekundären Maßen durchgeführt (siehe 8.3.6). In den Kategorien der Expositionsverfahren, der PCT und dem Vergleich TF vs. NTF ergab der Test auf Heterogenität keine signifikanten Ergebnisse.

Tabelle 12

Kategorienvergleiche zur relativen Effektivität von Bona Fide-Therapien zu Behandlungsende

Kategorie	Maße	k	Mittelwertsvergleiche			Heterogenität			
			ES	95% KI	<i>p</i>	<i>T</i>	<i>Q</i>	<i>p</i>	<i>I</i> ² (%)
EX	Prim.	7	0.17	[-0.04, 0.37]	.112	0	4.14	.658	0
	Sek.	7	-0.06	[-0.24, 0.13]	.556	0.002	6.22	.399	0.01
PE	Prim.	7	0.01	[-0.26, 0.28]	.927	0.250	14.98	.020	55.29
	Sek.	7	-0.02	[-0.39, 0.35]	.914	0.421	26.70	.0002	81.72
EXPE	Prim.	14	0.10	[-0.07, 0.26]	.240	0.168	19.33	.113	32.73
	Sek.	14	-0.02	[-0.20, 0.17]	.853	0.246	33.18	.002	56.59
CBT	Prim.	15	-0.02	[-0.17, 0.14]	.843	0.173	21.89	.081	38.05
	Sek.	15	-0.03	[-0.18, 0.13]	.735	0.203	26.25	.024	50.47
EMDR	Prim.	6	0.01	[-0.45, 0.48]	.959	0.483	17.02	.005	70.37
	Sek.	6	0.17	[-0.32, 0.66]	.505	0.513	17.97	.003	72.77
PCT	Prim.	6	-0.17	[-0.32, 0.01]	.038	0.077	4.38	.496	14.66
	Sek.	6	-0.04	[-0.15, 0.07]	.468	0.002	5.39	.370	0.02
TF vs. NTF	Prim.	10	0.05	[-0.12, 0.22]	.542	0.159	13.33	.148	40.76
	Sek.	10	0.05	[-0.05, 0.14]	.361	0.001	10.17	.337	0

Anmerkungen. Eine positive Effektgröße bedeutet, dass diese Kategorie wirksamer ist als die Vergleichsverfahren. k = Anzahl der Vergleiche; ES = Effektgröße Hedges *g*; KI = Konfidenzintervall; *T* = Schätzwert der Standardabweichung der wahren Effektgrößen; *Q* = χ^2 -verteilte Prüfgröße für den Test auf Heterogenität; *I*² = Heterogenitätsmaß in Prozenten; EX = Expositionsverfahren; PE = Prolonged Exposure; EXPE = Expositionsverfahren und Prolonged Exposure; CBT = Kognitive Verhaltenstherapie; EMDR = Eye Movement Desensitization and Reprocessing; PCT = Gegenwartszentrierte Therapie; TF vs. NTF = traumafokussierte Behandlungsverfahren im Vergleich zu nicht-traumafokussierten Behandlungsverfahren; Prim. = Primäre PTSD-Maße (Symptome der PTSD); Sek. = Sekundäre Maße (Komorbide Symptomatik, traumabezogene Symptomatik und funktionsbezogene Maße).

Die Ergebnisse dieser Kategorienvergleiche stimmen zum Teil mit den Ergebnissen anderer Metaanalysen überein:

- **Expositionsverfahren:** In Vergleichen von Expositionsverfahren mit CBT oder EMDR wurden keine signifikanten Unterschiede zwischen den Verfahren festgestellt (Bradley et al., 2005; Davidson & Parker, 2001). In der vorliegenden Arbeit waren die Vergleichsverfahren CBT, EMDR, PCT,

Kognitive Therapie (CT), Hypnotherapie und ein anderes Expositionsverfahren.

- **PE:** Eine Metaanalyse zur Wirksamkeit von PE fand keinen signifikanten Gesamtunterschied zwischen PE und den Behandlungsverfahren EMDR, CT, dem Stressimpfungstraining und einem weiteren kognitiven Verfahren (Powers et al., 2010). Fünf (von sieben) Studien sind mit dieser Metaanalyse ident. In der vorliegenden Arbeit war zusätzlich noch PCT unter den Vergleichsbedingungen.
- **CBT:** Äquivalente Ergebnisse wurden für einen metaanalytischen Vergleich von CBT und anderen spezifischen Verfahren gefunden (Bradley et al., 2005). TFCBT im Vergleich mit EMDR brachte ebenfalls vergleichbare Ergebnisse (Bisson et al., 2007; Seidler & Wagner, 2006). In der vorliegenden Arbeit waren die Vergleichsverfahren PE, EMDR, PCT, Expositionsverfahren und zwei andere spezifische CBT-Verfahren.
- **EMDR:** Metaanalysen zeigten, dass EMDR und TFCBT beziehungsweise EMDR und verhaltenstherapeutische Verfahren vergleichbar effektiv sind (Bisson et al., 2007; Bradley et al., 2005; Van Etten & Taylor, 1998). Eine Metaanalyse mit fast identem Studienkorpus konnte keinen Unterschied zwischen den beiden Verfahren feststellen und fand als Ergebnis ebenfalls eine signifikant hohe Heterogenität. Aufgrund der geringen Studienanzahl wurden keine Moderatorenanalysen durchgeführt (Seidler & Wagner, 2006).

Die beiden Ergebnisse jedoch, zu denen es keine übereinstimmenden Befunde gibt, und die daher einer eingehenden Diskussion bedürfen (siehe 9.2.1 und 9.2.2), sind die geringere Wirksamkeit der PCT, ein Ergebnis, das nicht der Schlussfolgerung von Benish et al. (2008) entspricht, dass alle Bona Fide-Psychotherapien gleich effektiv sind, und die äquivalenten Ergebnisse von traumafokussierten und nicht-traumafokussierten Behandlungsverfahren, die der aktuellen Evidenzlage widersprechen (Bisson et al., 2007; Ehlers et al., 2010).

In der vorliegenden Arbeit wurden auch Kategorienvergleiche zu zwei Follow-up-Zeitpunkten durchgeführt (siehe Tabelle 13).

Tabelle 13

Kategorienvergleiche zur relativen Effektivität von Bona Fide-Therapien: Follow-up Daten

Kategorien	Maße	k	Mittelwertsvergleiche				Heterogenität		
			ES	95% KI	<i>p</i>	<i>T</i>	<i>Q</i>	<i>p</i>	<i>I</i> ² (%)
zum ersten Follow-up									
EX	Prim.	7	-0.04	[-0.25, 0.17]	.705	0	2.88	.824	0
	Sek.	7	-0.06	[-0.25, 0.13]	.506	0.002	5.96	.428	0
PE	Prim.	7	0.10	[-0.14, 0.34]	.406	0.188	9.43	.151	38.77
	Sek.	7	0.18	[-0.12, 0.47]	.240	0.288	14.06	.029	62.22
EXPE	Prim.	14	0.04	[-0.11, 0.20]	.575	0.123	13.67	.398	19.90
	Sek.	14	0.06	[-0.11, 0.22]	.515	0.182	21.89	.057	38.84
CBT	Prim.	14	0.01	[-0.17, 0.20]	.906	0.224	24.64	.026	48.50
	Sek.	14	-0.02	[-0.16, 0.12]	.739	0.133	20.95	.074	27.78
EMDR	Prim.	5	-0.02	[-0.62, 0.57]	.940	0.563	13.32	.010	71.64
	Sek.	5	-0.05	[-0.66, 0.56]	.874	0.581	13.65	.009	72.66
PCT	Prim.	6	-0.17	[-0.30, 0.04]	.010	0	3.71	.592	0
	Sek.	6	-0.05	[-0.18, 0.09]	.477	0.063	7.26	.202	12.92
TF vs. NTF	Prim.	10	0.11	[-0.04, 0.26]	.160	0.110	10.72	.296	23.34
	Sek.	10	0.03	[-0.09, 0.14]	.656	0.056	13.64	.136	9.38
zum zweiten Follow-up									
CBT	Prim.	8	0.02	[-0.40, 0.44]	.926	0.514	25.14	.001	80.21
	Sek.	8	-0.06	[-0.37, 0.24]	.680	0.362	23.12	.002	72.45
TF vs. NTF	Prim.	7	0.17	[-0.15, 0.48]	.298	0.327	15.74	.015	68.83
	Sek.	7	0.03	[-0.09, 0.16]	.602	0.001	6.04	.419	0

Anmerkungen. Eine positive Effektgröße bedeutet, dass diese Kategorie wirksamer ist als die Vergleichsverfahren. *k* = Anzahl der Vergleiche; ES = Effektgröße Hedges *g*; KI = Konfidenzintervall; *T* = Schätzwert der Standardabweichung der wahren Effektgrößen; *Q* = χ^2 -verteilte Prüfgröße für den Test auf Heterogenität; *I*² = Heterogenitätsmaß in Prozenten; EX = Expositionsverfahren; PE = Prolonged Exposure; EXPE = Expositionsverfahren und Prolonged Exposure; CBT = Kognitive Verhaltenstherapie; EMDR = Eye Movement Desensitization and Reprocessing; PCT = Gegenwartszentrierte Therapie; TF vs. NTF = traumafokussierte Behandlungsverfahren im Vergleich zu nicht-traumafokussierten Behandlungsverfahren; Prim. = Primäre PTSD-Maße (Symptome der PTSD); Sek. = Sekundäre Maße (Komorbid Symptomatik, traumabezogene Symptomatik und funktionsbezogene Maße).

Obwohl es wünschenswert wäre, Daten von längeren Nachbeobachtungszeiträumen miteinander zu vergleichen, ist festzuhalten, dass sich nach dem ersten Follow-up nach durchschnittlich 4.67 Monaten ein durchaus ähnliches Ergebnismuster zeigte wie zu Behandlungsende. Auch die Anzahl der Vergleiche, auf denen die Analysen beruhten, stimmten fast überein, da nur eine Studie mit einem Vergleich von CBT und EMDR keine Follow-up-Daten für die primären und sekundären Maße berichtete (Power et al., 2002).

In neun Studien war auch ein zweites Follow-up durchgeführt worden. Damit war es möglich, die Daten für zwei Kategorien von Behandlungsverfahren für diesen Erhebungszeitpunkt auszuwerten. Die CBT erwies sich als gleich wirksam wie die Vergleichsverfahren, und auch für den Vergleich TF vs. NTF konnte kein signifikanter Unterschied festgestellt werden. Nur die hohe Heterogenität war in beiden Kategorien auffällig (siehe 8.2).

8.3.2 Detailanalysen zu den primären und sekundären Maßen

Die primären PTSD-Maße wurden auch getrennt nach Interview- und Selbstbeurteilungsverfahren ausgewertet, die sekundären Maße getrennt nach störungsbezogenen und funktionsbezogenen Maßen. In den Detailanalysen zeigten sich folgende Unterschiede zwischen diesen Variablen:

Interviews versus Selbstbeurteilungsfragebögen

Im Vergleich TF vs. NTF fand sich ein signifikanter mittlerer Effekt. Die traumafokussierten Verfahren waren zum zweiten Follow-up signifikant wirksamer, wenn die Erhebung nur mit Selbstbeurteilungsfragebögen erfolgt war ($g = 0.24$, $p = .033$). Basis waren hier fünf Vergleiche. Dieser Trend zeichnete sich auch zu den anderen Erhebungszeitpunkten ab. Zum Behandlungsende war $g = 0.15$ ($p = .118$), zum ersten Follow-up war $g = 0.12$ ($p = .070$). Hier wurden jeweils sechs Vergleiche in die Analyse einbezogen. Der Test auf Heterogenität war in allen drei Fällen nicht signifikant.

Im Gegensatz dazu war in der Kategorie der Expositionsverfahren der mittlere Effekt bei der Erhebung mit Interviews größer und fast signifikant ($g = 0.23$, $p = .109$, 95% KI = [-0.05, 0.52]) als bei der Verwendung von Selbstbeurteilungsfragebögen ($g = 0.06$, $p = .627$, 95% KI = [-0.19, 0.31]). Es überschritten sich allerdings die Konfi-

denzintervalle. Sowohl im Bereich PTSD als auch im Bereich Depression wurden größere Effekte für Interviewverfahren im Vergleich zu Selbstbeurteilungsverfahren gefunden (Cuijpers et al., 2010; Taylor und Harvey, 2009). Diese Detailanalyse konnte nur noch für die Kategorien EMDR und CBT durchgeführt werden. Dabei ließ sich kein eindeutiger Trend erkennen, dass eines der beiden Messverfahren durchgängig größere Effektgrößen zeigen würde. Mitunter wichen die beiden Erhebungsverfahren deutlich voneinander ab, für EMDR war sogar die Richtung des Effekts je nach Erhebungsmethode verschieden. Dies könnte als weiterer Hinweis dafür gedeutet werden, dass das Aggregieren verschiedener Effektgrößen, diesmal von unterschiedlichen Messinstrumenten, möglicherweise vorhandene Unterschiede zwischen den Behandlungsverfahren überdeckt.

Die Messverfahren können sich aber auch hinsichtlich der Heterogenität unterscheiden. Hier fand sich in den Analysen ein Hinweis darauf, dass, wenn ein Unterschied zwischen den Messverfahren hinsichtlich des Ausmaßes an Heterogenität vorhanden war, die Erhebungen mittels Selbstbeurteilungsfragebögen größere I^2 -Werte zu Behandlungsende aufwiesen, während zum ersten Follow-up die größeren Werte an Heterogenität bei Erhebungen mittels Interviews gefunden wurden.

Störungsbezogene versus funktionsbezogene sekundäre Maße

Es war nur in vier Kategorien von Behandlungsverfahren möglich, diese Detailanalysen durchzuführen. Für die PCT war der mittlere Effekt für die störungsbezogenen Maße sowohl zum Behandlungsende ($g = -0.11, p = .001$) signifikant als auch zum ersten Follow-up ($g = -0.15, p = .063$) nahe der Signifikanzgrenze, während in den funktionsbezogenen Maßen wie beim Upper Bound-Verfahren kein Unterschied zwischen den Therapien sichtbar wurde. Wenn beide Maße zusammengefasst wurden, waren die sekundären Maße deutlich nicht signifikant. Im Vergleich TF vs. NTF wiederholte sich dieses Bild, wenn auch nicht so ausgeprägt. Dabei ist zu berücksichtigen, dass die PCT einen großen Teil dieser Kategorie ausmachten. In den Kategorien CBT und EXPE zeigte sich dieses Muster nicht, sodass eine mögliche Erklärung darin liegen könnte, dass non-direktive Therapieverfahren in den funktionsbezogenen Maßen gleich effektiv, in den störungsbezogenen sekundären Maßen allerdings weniger effektiv sind wie andere Behandlungsverfahren.

8.3.3 Sensitivitätsanalysen zu den primären und sekundären Maßen

Es wurden drei verschiedene Arten von Sensitivitätsanalysen durchgeführt, um die Ergebnisse auf ihre Robustheit zu überprüfen (siehe 6.5).

Abhängigkeiten: Grund für die Sensitivitätsanalyse war, dass in der Kategorie der Expositionsverfahren zwei Vergleiche enthalten waren, die von einer Studie stammten (Brom et al., 1989), und eine Studie mit einem Vergleich zweier Expositionsverfahren inkludiert war (Bryant et al., 2008). Auch in der Kategorie CBT wurde in zwei Studien ein CBT-Verfahren mit einem anderen CBT-Verfahren verglichen (Cook et al., 2010; Hien et al., 2004). In den entsprechenden Kategorien wurden diese Vergleiche nacheinander ausgeschlossen und die Metaanalysen ohne diese Effekte gerechnet. Dies änderte jedoch nichts an den oben berichteten Ergebnissen, die hinsichtlich der Abhängigkeiten als robust bezeichnet werden können.

Publication Bias: Das Trim-and-fill-Verfahren (Duval & Tweedie, 2000) wurde für jede Kategorie der Behandlungsverfahren für den Zeitpunkt des Behandlungsendes angewandt. Damit ist es möglich festzustellen, ob Studien in der Analyse fehlen, und wie groß die unverzerrte Effektgröße unter Berücksichtigung dieser fehlenden Studien ist. In den meisten Fällen wurden keine fehlenden Studien angegeben, beziehungsweise änderte die Berechnung der unverzerrten Effektgröße nichts Wesentliches am Ergebnis. Erwähnenswert ist aber, dass für die PCT nach Berücksichtigung einer fehlenden Studie die neue Effektgröße mit -0.20 ($p = .011$) angegeben wurde. Nach Cohen (1988) bedeutet eine Effektgröße von 0.20 bereits einen geringen Effekt. Auch unter Weglassung eines einflussreichen Falls, nämlich der Studie von Schnurr et al. (2007), wodurch der mittlere Effekt von vornherein seine Signifikanz verlor (siehe unten), blieb der mittlere Effekt auch nach Berücksichtigung des Publication Bias nicht signifikant. Beides sind also in Bezug auf den Publication Bias robuste Ergebnisse. Die bereits erwähnte auffällig hohe Effektgröße der primären PTSD-Maße der Expositionsverfahren reduzierte sich nach Berücksichtigung zweier fehlender Studien auf $g = 0.11$ ($p = .280$). Zusammenfassend lässt sich festhalten, dass die Analysen zum Publication Bias keine Änderungen brachten.

Einflussreiche Fälle: Die Analyse der einflussreichen Fälle wurde mit dem Leave-One-Out-Verfahren durchgeführt (siehe 6.5.3). Für die Expositionsverfahren und EMDR waren die Ergebnisse sowohl für die primären als auch für die sekundären

Maße robust. Das bedeutet, dass, egal welche Studie ausgeschlossen wurde, sich weder etwas an der Signifikanz der Effektgrößen noch am Ergebnis des Tests auf Heterogenität änderte.

- **PE:** Zwei Studien erwiesen sich hinsichtlich der Heterogenität der primären PTSD-Maße als einflussreich (Resick, Nishith, Weaver, Astin, & Feurer, 2002; Schnurr et al., 2007). Jede dieser Studien brachte für sich so viel Heterogenität in die Analyse ein, dass der Q -Wert dadurch signifikant wurde. Würde man auch nur eine dieser beiden Studien aus der Analyse ausschließen, wäre keine signifikante Heterogenität für die Kategorie der PE mehr vorhanden.
- **EXPE:** In dieser Kategorie zeigte sich auch der Einfluss der Resick et al. Studie. Durch ihren Ausschluss ergab sich mit $g = 0.19$ ($p = .005$) ein signifikanter Unterschied zu Gunsten von EXPE, wahrscheinlich aufgrund der größeren Teststärke (13 Studien) dieser Kategorie. Die Heterogenität ging dann mit $I^2 = 2.20\%$ auch gegen Null.
- **CBT:** Die Ergebnisse waren robust für die primären PTSD-Maße, allerdings brachten drei Studien beträchtliche Heterogenität in die sekundären Maße ein (Foa et al., 1999; Power et al., 2002; Resick et al., 2002). Wenn jede dieser Studien in der Analyse fehlen würde, wäre der Test auf Heterogenität nicht mehr signifikant.
- **PCT:** Zwei Studien zeigten einen bedeutsamen Einfluss. Einmal die Studie, die von Wampold et al. (2010) als nicht-bona fide qualifiziert wurde (Schnurr et al., 2007), und zum anderen die Studie der gleichen ForscherInnengruppe, in der zwei Gruppenverfahren untersucht wurden (Schnurr et al., 2003). Würde jede dieser Studien für sich genommen aus der Analyse ausgeschlossen werden, wäre der statistisch bedeutsame mittlere Effekt nicht mehr signifikant.
- **TF vs. NTF:** In diesem Vergleich gab es ebenfalls eine einflussreiche Studie (Hien et al., 2004). Wurde diese in der Analyse nicht berücksichtigt, so zeigte sich ein signifikanter Unterschied zu Gunsten der traumafokussierten Behandlungsverfahren für die primären PTSD-Maße ($g = 0.14$, $p = .030$).

- Um explizit zu überprüfen, ob die Studie von Schnurr et al. (2007) auch einen Einfluss auf die anderen relevanten Kategorien hatte (PE, EXPE, TF vs. NTF), wurden alle betreffenden Metaanalysen nochmals ohne diese Studie gerechnet. In allen Kategorien außer PCT blieben die Ergebnisse im Wesentlichen gleich. Die PCT waren ohne diese Studie als Kategorie gleich wirksam wie die Vergleichsverfahren ($g = -0.09$, $p = .289$ zu Behandlungs-ende; $g = -0.13$, $p = .122$ zum ersten Follow-up).

Zusammenfassend ist festzuhalten, dass fast alle einflussreichen Studien als Gemeinsamkeiten ITT-Analysen und eine große Anzahl von PatientInnen aufwiesen (Hien et al., 2004; Resick et al., 2002; Schnurr et al., 2003; Schnurr et al., 2007). Insofern ist es nachvollziehbar, dass bei einer geringen Studienanzahl die Auswirkungen aufgrund des Ausschlusses einer solchen Studie beträchtlich sind.

8.3.4 Klinische Signifikanz

Bei dieser Variable handelt es sich um die Genesungsraten (keine PTSD mehr diagnostizierbar) beziehungsweise, für den Fall, dass keine Informationen zur Genesungsrate vorlagen, um die Besserungsraten, so wie diese von den jeweiligen StudienautorInnen definiert wurden. In den einzelnen Kategorien von Behandlungsv erfahren wurden von 50% bis 71% der inkludierten Studien Genesungsraten berichtet.

Zur Untersuchung der klinischen Signifikanz wurde das relative Risiko (*RR*) berechnet. Ein *RR* von 1 bedeutet, dass kein Unterschied zwischen den Verfahren besteht, ein *RR* über 1 bedeutet, dass die Kategorie von Verfahren ein größeres Risiko hat, in diesem Zusammenhang besser als Chance ausgedrückt, ein klinisch signifikantes Ergebnis, also höhere Genesungs- beziehungsweise Besserungsraten zu erzielen. In diesem Sinne wäre die Kategorie von Verfahren wirksamer als die Vergleichsverfahren. Ein *RR* unter 1 bedeutet, dass die Kategorie von Verfahren eine geringere Chance (Risiko) hat, ein klinisch signifikantes Ergebnis zu erzielen, also weniger wirksam ist als die Vergleichsverfahren.

Zur besseren Interpretation der Ergebnisse wurden Schwellenwerte herangezogen, die für Studien geeignet sind, in denen zwei aktive Verfahren miteinander verglichen werden. Diese Schwellenwerte, die klinisch bedeutsame Ergebnisse anzeigen, wurden von der *NICE Guideline Development Group* erarbeitet (Bisson et al., 2007).

Ein *RR* von 0.80 oder weniger beziehungsweise ein *RR* von 1.25 oder mehr wurden als klinisch bedeutsam erachtet. Da nur wenige Studien Follow-up-Daten zur klinischen Signifikanz berichtet hatten, wurde nur ein Nacherhebungszeitpunkt definiert. Von jenen Studien, die zwei Follow-ups durchführten, wurde der spätere Zeitpunkt für die Analyse herangezogen.

Wie in Tabelle 14 ersichtlich, waren die meisten Kategorien von Verfahren hinsichtlich der klinischen Signifikanz gleich wirksam wie die Vergleichsverfahren. Nur im Vergleich TF vs. NTF erzielten die traumafokussierten Verfahren klinisch bedeutsam höhere Genesungs- und Besserungsraten, wobei man einschränkend festhalten muss, dass das Ergebnis nicht signifikant war. In diesem Zusammenhang ist es interessant zu erwähnen, dass die PCT, die einen Großteil der nicht-traumafokussierten Verfahren ausmachten, zum Zeitpunkt des Behandlungsendes nicht analysiert werden konnten, da in einer Studie die klinische Signifikanz nur zum Follow-up erhoben worden war (Neuner et al., 2004). Zum Follow-up erzielte diese Kategorie klinisch bedeutsam geringere Genesungs- und Besserungsraten, ohne dass das Ergebnis allerdings signifikant war ($RR = 0.79, p = 0.76$). Im Vergleich TF vs. NTF waren alle Therapieverfahren zum Follow-up wieder gleich wirksam, die traumafokussierten Verfahren lagen aber nahe an der Signifikanzgrenze ($RR = 1.21, p = .076$). Man kann also davon sprechen, dass es einen Trend in Richtung höhere Genesungsraten für traumafokussierte Verfahren gibt.

Der Test auf Heterogenität war wie bei den primären und sekundären Maßen für EMDR und PE (und daher auch für EXPE) signifikant, die Heterogenität lag mit I^2 -Werten von 67.01% für EMDR und 66.76% für PE etwas über dem mittleren Bereich. Leider waren zu wenige Studien für eine Moderatorenanalyse vorhanden.

Die aktuelle Datenlage lässt den Schluss zu, dass die inkludierten Therapieverfahren hinsichtlich ihrer Genesungs- und Besserungsraten sowohl zu Behandlungsende als auch zum Follow-up vergleichbar wirksam sind. Ob der Trend zu Gunsten traumafokussierter Therapieverfahren als robust bezeichnet werden kann, wurde in den Sensitivitätsanalysen untersucht.

Tabelle 14

Relatives Risiko zur klinischen Signifikanz zum Behandlungsende

Kategorie	k	Gen.	Mittelwertsvergleich			Heterogenität			
			RR	95% KI	p	T	Q	p	I ² (%)
PE	7	5	1.13	[0.84, 1.52]	.415	0.311	16.05	.014	66.76
EXPE	9	6	1.13	[0.89, 1.43]	.312	0.264	18.58	.017	57.60
CBT	11	7	0.95	[0.83, 1.08]	.453	0	11.47	.322	0
EMDR	6	3	1.06	[0.76, 1.48]	.733	0.322	14.51	.013	67.01
TF vs. NTF	6	4	1.31	[0.95, 1.79]	.095	0.251	9.69	.084	46.32

Anmerkungen. Ein *RR* (Relatives Risiko) von 1 bedeutet, dass das Verfahren in etwa die gleichen Genesungs- bzw. Besserungsraten hat wie die Vergleichsverfahren, ein *RR* über 1.25 gilt als klinisch bedeutsamer Unterschied zu Gunsten dieser Kategorie, ein *RR* unter 0.80 zu Lasten dieser Kategorie. k = Anzahl der Vergleiche; Gen. = Anzahl der Studien, die Genesungsraten berichteten; KI = Konfidenzintervall; T = Schätzwert der Standardabweichung der wahren Effektgrößen; Q = χ^2 -verteilte Prüfgröße für den Test auf Heterogenität; I² = Heterogenitätsmaß in Prozenten; PE = Prolonged Exposure; EXPE = Expositionsverfahren und Prolonged Exposure; CBT = Kognitive Verhaltenstherapie; EMDR = Eye Movement Desensitization and Reprocessing; TF vs. NTF = traumafokussierte Behandlungsverfahren im Vergleich zu nicht-traumafokussierten Behandlungsverfahren.

Sensitivitätsanalysen

- **Abhängigkeiten:** In den Daten zur klinischen Signifikanz waren keine Abhängigkeiten vorhanden.
- **Publication Bias:** Ein bedeutsames Ergebnis war, dass für die Kategorie PE nach Berücksichtigung zweier fehlender Studien das *RR* von 1.13 auf 1.30 ($p = .164$) anstieg, was nach den Schwellenwerten ein klinisch bedeutsames Ergebnis zu Gunsten der PE darstellt. Ein ähnliches Bild ergab sich auch für die Kategorie EXPE mit einem Anstieg des *RR* von 1.13 auf 1.25 ($p = .132$).
- **Einflussreiche Fälle:** Eine einflussreiche Studie (Ironson et al., 2002) betraf die Kategorien PE und EXPE. Deren Ausschluss bewirkte, dass sowohl für EXPE ($RR = 1.22$, $p = .043$) als auch für PE ($RR = 1.26$, $p = .035$) ein signifikanter Unterschied zwischen den Interventionen vorhanden war, und dass der Test auf Heterogenität nicht mehr signifikant war. Als einflussreich in

Bezug auf die Heterogenität erwies sich die Studie von Schnurr et al. (2007), da sich durch ihren Ausschluss diese wesentlich reduzierte (EXPE: $I^2 = 13.97$, PE: $I^2 = 18.66\%$). Die Ergebnisse für CBT waren robust, aber der Ausschluss der Studie von Schnurr et al. (2003) würde ein signifikantes Ergebnis zu Gunsten der traumafokussierten Verfahren bringen ($RR = 1.46$, $p = .034$), da dann die verbleibende Schnurr et al. (2007) Studie wesentlich ins Gewicht fallen würde. Würde man diese Studie ausschließen, wären die traumafokussierten Behandlungsverfahren eindeutig gleich wirksam wie die nicht-traumafokussierten Verfahren, ebenso wie die Kategorien PE und EXPE nicht mehr signifikant wirksamer wären (trotz der Ironson et al. Studie).

Zusammenfassend ist festzustellen, dass aufgrund der geringen Studienanzahl eine Studie (Schnurr et al., 2007) einen großen Einfluss auf die betreffenden Kategorien (PE, EXPE, PCT, TF vs. NTF) hatte. Da diese Studie hinsichtlich der Bona Fide-Definition umstritten ist, könnte dies als Hinweis darauf gedeutet werden, dass die unterschiedlichen Ergebnisse von BefürworterInnen und GegnerInnen des Dodo Bird Verdicts tatsächlich (auch) auf das Anwenden beziehungsweise Nicht-Anwenden der Bona Fide-Definition zurückzuführen sind, wie Wampold et al. (2010) postulierten. Diese Argumentation wird unter 9.2.3 nochmals ausgeführt.

Eine zusätzliche Erklärung ist, dass gerade, wenn die Studienanzahl gering ist, große Studien wie die beiden Schnurr et al. Studien (325 und 284 Personen) besonders ins Gewicht fallen.

8.3.5 Dropouts

Das Risiko für Dropouts wurde ebenfalls mit dem RR berechnet. Wie Tabelle 15 zu entnehmen ist, lag PE an der Schwelle zu einem klinisch bedeutsamen Ergebnis. Das Risiko für Dropouts war für PE höher als für die Vergleichsverfahren, für PCT hingegen niedriger. Beide Ergebnisse waren nicht signifikant. In den Kategorien PCT, CBT und dem Vergleich TF vs. NTF war die Heterogenität signifikant. Für die beiden letzteren wurden Moderatorenanalysen durchgeführt.

Tabelle 15

Relatives Risiko für Dropouts bis zum Behandlungsende

Kategorie	k	Mittelwertsvergleich			<i>T</i>	Heterogenität		
		<i>RR</i>	95% KI	<i>p</i>		<i>Q</i>	<i>p</i>	<i>I</i> ² (%)
EX	7	1.12	[0.69-1.79]	.666	0	3.01	.808	0
PE	7	1.25	[0.76-2.03]	.380	0.377	11.19	.083	39.03
EXPE	14	1.22	[0.90-1.67]	.201	0.228	14.91	.313	16.07
CBT	15	1.20	[0.86-1.68]	.281	0.390	24.17	.044	42.92
EMDR	6	0.83	[0.53-1.28]	.397	0	5.27	.384	0
PCT	6	0.77	[0.33-1.80]	.554	0.841	14.56	.012	82.19
TF vs. NTF	10	0.97	[0.56-1.70]	.924	0.695	31.18	.0003	76.35

Anmerkungen. Ein *RR* (Relatives Risiko) von 1 bedeutet, dass die Verfahren das gleiche Risiko für Dropouts haben, ein *RR* über 1.25 gilt als klinisch bedeutsam größeres Risiko, ein *RR* unter 0.80 als geringeres Risiko. *k* = Anzahl der Vergleiche; KI = Konfidenzintervall; *T* = Schätzwert der Standardabweichung der wahren Effektgrößen; *Q* = χ^2 -verteilte Prüfgröße für den Test auf Heterogenität; *I*² = Heterogenitätsmaß in Prozenten; EX = Expositionsverfahren; PE = Prolonged Exposure; EXPE = Expositionsverfahren und PE; CBT = Kognitive Verhaltenstherapie; EMDR = Eye Movement Desensitization and Reprocessing; PCT = Gegenwartszentrierte Therapie; TF vs. NTF = Traumafokussierte vs. nicht-traumafokussierte Therapien.

Sensitivitätsanalysen:

- **Abhängigkeiten:** Die durchgeführten Sensitivitätsanalysen zeigten, dass die Ergebnisse robust sind.
- **Publication Bias:** Nach Berücksichtigung dreier fehlender Studien ergab sich für die Expositionsverfahren ein *RR* von 1.30 (*p* = .238) und damit ein klinisch bedeutsam höheres Risiko für Dropouts. In der Kategorie PE wurde eine fehlende Studie berücksichtigt. Die neue Effektgröße war *RR* = 1.15 (*p* = .608), es war kein Unterschied mehr zwischen den Verfahren vorhanden. Für die Kategorie EXPE konnte das Trim-and-fill-Verfahren leider nicht durchgeführt werden. Alle anderen Ergebnisse waren hinsichtlich eines möglichen Publication Bias robust.
- **Einflussreiche Fälle:** Nur in zwei Kategorien von Behandlungsverfahren verursachte der Ausschluss einer Studie (Cottraux et al., 2008), in der die

Ergebnisse gegen den Trend der meisten anderen Studien dieser Kategorien gerichtet waren, bedeutsame Ergebnisänderungen. Sowohl für CBT als auch für PCT war die Effektgröße danach klinisch bedeutsam und signifikant (CBT: $RR = 1.38$, $p = .005$; PCT: $RR = 0.57$, $p = .000$), was bedeutet, dass das Risiko eines Dropouts bei CBT-Verfahren größer ist und bei PCT-Verfahren geringer. Allerdings ist diese Studie auch insofern bedeutsam, als dass sie eine relativ große Stichprobe umfasste.

Zusammenfassend ist festzuhalten, dass die Daten einen Trend nahelegen, dass Behandlungsverfahren der Kategorien EXPE und CBT ein höheres Risiko für Dropouts haben und die PCT ein geringeres Risiko.

8.3.6 Moderatorenanalysen

Aufgrund der geringen Studienanzahl in den Kategorien EMDR und PE bestand nur die Möglichkeit, für die Kategorien CBT und EXPE und für den Vergleich TF vs. NTF Moderatorenanalysen durchzuführen. Innerhalb dieser Kategorien gab es nur wenige Analysen, in denen signifikante Heterogenität vorhanden war. Für EXPE wurden für die sekundären Maße zu Behandlungsende alle Moderatoren überprüft (siehe 6.6), aber keiner konnte dazu beigetragen, einen signifikanten Anteil an Heterogenität aufzuklären. Für die sekundären Maße der CBT zum Behandlungsende und für die primären PTSD-Maße zum ersten Follow-up zeigten Sensitivitätsanalysen, dass einzelne Studien beträchtliche Heterogenität mit sich brachten (siehe 8.3.3). Dennoch wurden auch diese Analysen auf Moderatoren überprüft. In beiden Fällen zeigte sich das Qualitätsrating nach Jadad et al. (1996) als signifikanter Moderator, um die Heterogenität aufzuklären. Dabei wurde ersichtlich, dass eine hohe Studienqualität deutlich geringere Unterschiede zwischen den Behandlungsverfahren mit sich brachte als Studien, die die Qualitätskriterien nicht ausreichend erfüllten (siehe Tabelle 16).

Einige Vergleiche wurden unter Berücksichtigung des Moderators sogar signifikant, was bedeutet, dass auf dieser Qualitätsstufe ein Unterschied zwischen den Therapien vorhanden war. Dies galt zumindest für die CBT-Verfahren, da das Rating in der Kategorie EXPE und in den Moderatorenanalysen der Dropouts (CBT und TF vs. NTF) nicht signifikant war.

Tabelle 16

Moderatorenanalyse: Unterschiede in den Effektgrößen für CBT je nach Studienqualität

Qualitätsrating ^a	k	Primäre PTSD-Maße zum Follow-up			Sekundäre Maße zum Behandlungsende		
		ES	<i>p</i>	95% KI	ES	<i>p</i>	95% KI
0	2	-0.45	.006	[-0.77, -0.13]	-0.26	.196	[-0.66, 0.14]
1	3	0.31	.023	[0.04, 0.57]	0.43	.001	[0.18, 0.67]
2	4	-0.09	.630	[-0.46, 0.28]	-0.21	.108	[-0.46, 0.05]
3	1	0.09	.672	[-0.33, 0.50]	-0.03	.882	[-0.44, 0.38]
4	4/5 ^b	0.09	.252	[-0.07, 0.25]	-0.02	.744	[-0.15, 0.11]

Anmerkungen. Eine positive Effektgröße bedeutet, dass CBT-Verfahren wirksamer sind als die Vergleichsverfahren. CBT = Kognitive Verhaltenstherapie; k = Anzahl der inkludierten Studien; ES = Effektgröße Hedges *g*; KI = Konfidenzintervall; Primäre PTSD-Maße = Symptome der PTSD; Sekundäre Maße = Komorbide Symptomatik, traumabezogene Symptomatik und funktionsbezogene Maße.

^anach Jadad et al. (1996) werden Punkte für das Einhalten methodischer Kriterien vergeben.

^b4 Studien zum Follow-up, 5 zum Behandlungsende.

Da die Teststärke aufgrund der geringen Studienanzahl eher gering war, sind die signifikanten Ergebnisse umso bedeutsamer. Es darf aber nicht geschlossen werden, dass die anderen Moderatorvariablen keinen Effekt haben. Es könnte sein, dass die Effekte sehr klein sind oder die Teststärke zu gering war, um einen signifikanten Effekt zu entdecken (Borenstein et al., 2009).

Für das Studienqualitätsrating nach Jadad et al. (1996) lässt sich schließen, dass es geeignet ist, auch im Bereich der Psychotherapieforschung zwischen Studien nach methodischer Qualität zu differenzieren. Gleichzeitig kann das Ergebnis als Hinweis darauf interpretiert werden, dass Schlüsse hinsichtlich der relativen Effektivität von Bona Fide-Psychotherapien mit Vorsicht zu ziehen sind, wenn nur Wirksamkeitsstudien von geringer methodischer Qualität vorliegen.

Für die sekundären Maße der CBT zu Behandlungsende wurde auch der *Q*-Test für den Moderator *Zeit nach dem Trauma* signifikant ($QM = 9.09$, $p = .003$) und konnte die Heterogenität aufklären ($Q = 7.24$, $p = .512$). Leider konnte dieser Moderator für die beiden anderen Analysen aufgrund zu weniger Daten nicht durchgeführt werden.

9 Zusammenfassung und Diskussion

Die Haupteergebnisse werden nun nochmals zusammengefasst und zu den Forschungsfragen der vorliegenden Arbeit in Beziehung gesetzt. Dabei sollen folgende Fragestellungen beantwortet werden: Konnten Unterschiede in den Ergebnissen je nach metaanalytischer Methodik beobachtet werden, und bringt eine möglichst differenzierende Strategie Unterschiede zwischen Behandlungsverfahren zum Vorschein? Welche Therapierichtungen zeigen äquivalente Ergebnisse, und wo sind Unterschiede in der Effektivität zu finden? Gibt es Kategorien von Behandlungsverfahren, die signifikant wirksamer sind als die Vergleichsverfahren?

9.1 Zusammenfassung der Haupteergebnisse

Die Frage nach der relativen Wirksamkeit von Bona Fide-Psychotherapien für die Posttraumatische Belastungsstörung wurde mit Hilfe dreier verschiedener metaanalytischer Verfahren zu beantworten versucht.

Mittels des Random Signs-Verfahrens wurden sowohl der Studienkorpus von Benish et al. (2008) als auch der aktualisierte Studienkorpus analysiert, um festzustellen, ob die Effektgrößen gemäß der Nullhypothese homogen um Null verteilt sind. Alle Analysen konnten diese Annahme und damit die Ergebnisse von Benish et al. nicht bestätigen. Die in diesem Verfahren angewandte zufällige Vorzeichenzuteilung führte entweder zu instabilen Ergebnissen oder zu einer heterogenen Verteilung der Effektgrößen. Gründe dafür sind die von der Benish et al. Studie unterschiedlichen Effektgrößen, die aus einer anderen Variablenzusammensetzung, der Verwendung einer anderen Datenbasis in zwei Fällen und einem unterschiedlichen Umgang mit fehlenden Werten resultierten. Ein weiterer Grund war die größere Teststärke der vorliegenden Arbeit, die 24 anstelle von 17 Vergleichen analysierte. Drittens und methodisch von besonderer Bedeutung ist, dass in dieser Arbeit gezeigt wurde, dass das Random Signs-Verfahren mit nur einer einzigen zufälligen Vorzeichenvergabe ein unzuverlässiges Verfahren ist, dessen Ergebnis nicht als Beleg für die Nullhypothese gesehen werden sollte.

Legt man die Interpretation von Benish et al. den Ergebnissen dieser Arbeit zu Grunde, muss nach diesem Verfahren der Schluss gezogen werden, dass das Dodo Bird Verdict, dass alle Bona Fide-Psychotherapien gleich wirksam sind, widerlegt ist.

Das zweite Verfahren wird als Upper Bound-Verfahren bezeichnet (Benish et al., 2008; Wampold et al., 1997). Um eine Obergrenze eines möglichen Unterschiedes zwischen Therapieverfahren zu erhalten, wurde eine Metaanalyse mit den absoluten Werten der Effektgrößen, das heißt ohne Vorzeichen, durchgeführt. Das Ergebnis entsprach mit einem mittleren Effekt von $g = 0.23$ in etwa den Ergebnissen von Benish et al. (2008) und Wampold et al. (1997). Allerdings ist zu bezweifeln, dass dieses Ergebnis tatsächlich die Obergrenze eines mittleren Effekts darstellt. Einerseits lag die obere Grenze des Konfidenz- und Vorhersageintervalls bei 0.32 und andererseits konnte argumentativ dargelegt werden, dass dieser Effekt vielmehr ein Maß für die Variabilität in den Daten ist, da er sich notwendigerweise aus einer verzerrten Stichprobe errechnet.

Für sieben Kategorien von Behandlungsverfahren wurden Random-Effects-Metaanalysen nach dem Vorbild der Cuijpers et al. (2008) Studie zur vergleichenden Effektivität von Depressionsbehandlungen durchgeführt. Für die Expositionsverfahren, Prolonged Exposure, die Kognitive Verhaltenstherapie und EMDR bestätigten die Ergebnisse den aktuellen Stand der Forschung insofern, als dass keine Unterschiede im Vergleich zu anderen aktiven Verfahren gefunden wurden.

Zwei Ergebnisse sind hingegen von besonderer Bedeutung. Zum einen waren die gegenwartszentrierten Verfahren weniger effektiv als die Vergleichsverfahren. Allerdings hatte eine einzelne Studie auf dieses Ergebnis einen besonderen Einfluss (Schnurr et al., 2007). Als zweites bedeutsames Ergebnis erwiesen sich die traumafokussierten Behandlungsverfahren entgegen der derzeit gültigen Evidenzlage als gleich effektiv wie die nicht-traumafokussierten Verfahren. Beide Ergebnisse waren auch für das erste Follow-up gültig.

Detailanalysen zeigten, dass ein differenzierender Ansatz sehr wohl Unterschiede zwischen den Behandlungsverfahren sichtbar machen kann. Wurden nur Daten aus Selbstbeurteilungsfragebögen zur Analyse herangezogen, so zeigte sich ein Trend in Richtung einer höheren Wirksamkeit der traumafokussierten Verfahren, insbesondere in Hinblick auf längerfristige Wirkungseffekte. Ein ähnliches Bild ergab sich durch die Differenzierung von störungsbezogenen und funktionsbezogenen sekundären Maßen. Funktionsbezogene Maße maskierten einen Trend in Richtung

einer geringeren Wirksamkeit der gegenwartszentrierten Verfahren in den sekundären Maßen. Diese Befunde waren allerdings Einzelfälle. Der größte Teil der Detailanalysen veränderte nichts an den Ergebnissen.

Was die Klinische Signifikanz betrifft, so fanden sich auch hier kaum Unterschiede zwischen den Behandlungsverfahren. Ein Trend zu Gunsten traumafokussierter Therapieverfahren war für diese Variable vom Einfluss einer Studie abhängig (Schnurr et al., 2007). Die Analysen der Dropout-Raten der einzelnen Kategorien von Behandlungsverfahren konnten folgende Trends identifizieren: Expositionsverfahren inklusive Prolonged Exposure und Kognitive Verhaltenstherapie haben tendenziell ein höheres Risiko und gegenwartszentrierte Therapieverfahren ein niedrigeres Risiko für Dropouts.

Zur Aufklärung der hohen Heterogenität in einzelnen Analysen wurden Moderatorenanalysen durchgeführt, wobei nur zwei Moderatoren eine signifikante Reduktion an Heterogenität bewirkten. Das Studienqualitätsrating nach Jadad et al. (1996) war in der Lage, zwischen Studien der Kognitiven Verhaltenstherapie nach methodischer Qualität zu differenzieren und zeigte zumindest für diese Therapieform, dass die Therapien umso eher vergleichbar wirksam sind, je höher die Studienqualität ist. Die Zeit nach dem Trauma konnte als Moderator ebenfalls in einer Analyse die Heterogenität aufklären. Alle anderen Moderatoren zeigten keinen Effekt.

9.2 Sind Bona Fide-Psychotherapien gleich wirksam?

Da das Random Signs-Verfahren die Ergebnisse der Benish et al. (2008) Studie nicht bestätigen konnte, kann kein Omnibus-Test zur Beantwortung dieser Frage herangezogen werden. Betrachtet man die Ergebnisse aus den Kategorienvergleichen, so könnte man zusammenfassend antworten, dass viele Bona Fide-Psychotherapien gleich wirksam sind, aber nicht alle.

In den Kritiken des Dodo Bird Verdicts wird häufig damit argumentiert, dass die eine oder andere Studie aus diesen oder jenen Gründen aus einer Analyse ausscheiden oder in eine Analyse einbezogen werden sollte (Chambless, 2002; Ehlers et al., 2010). In dieser Arbeit konnte gezeigt werden, dass einzelne Studien für die Ergebnisse bestimmter Therapieformen tatsächlich ausschlaggebend sind, insbesondere wenn die

Studienanzahl gering ist und einzelne Studien mit einer großen Stichprobe und einer ITT-Analyse ein hohes Gewicht haben. Für andere Therapieverfahren waren die Ergebnisse jedoch stabil. Das spricht umso mehr dafür, in Metaanalysen Einschluss- und Kodierkriterien festzulegen, bevor Einzelstudien gesichtet werden und ihre Ergebnisse bekannt sind. Die Möglichkeit, dass eine Studienstichprobe einen Bias je nach theoretischer Ausrichtung der ForscherInnen enthält (Ehlers et al., 2010; Wampold et al., 2010), ist somit sowohl für die BefürworterInnen als auch für die KritikerInnen des Dodo Bird Verdicts gegeben. Das Kodierprotokoll im Voraus festzulegen und methodische Entscheidungen im Zweifelsfall zu diskutieren und schriftlich festzuhalten sind Maßnahmen um möglichen Verzerrungen entgegenzuwirken.

In der Benish et al. (2008) Studie führten die bereits erwähnten methodischen Entscheidungen bei der Effektgrößenberechnung zumindest teilweise zu niedrigeren Effektgrößen. Die Bona Fide-Definition als Einschlusskriterium in einer Metaanalyse zur Bewertung der relativen Effektivität zu nutzen, um therapeutisch intendierte Behandlungsverfahren selektieren zu können, scheint insbesondere für die Frage nach der Wirksamkeit von non-direktiven Verfahren ausschlaggebend zu sein.

Für die Posttraumatische Belastungsstörung ist aufgrund der vorliegenden Arbeit zu schließen, dass jene Therapieverfahren, die im engeren oder weiteren Sinne aus dem kognitiv-verhaltenstherapeutischen Bereich stammen, gleich wirksam sind. Ein differenzierender Ansatz kann zwar einzelne signifikante Ergebnisse sichtbar machen, das Gros der Analysen weist aber äquivalente Ergebnisse auf. Der Vorteil der vorliegenden Arbeit im Vergleich zu früheren Metaanalysen ist, dass als Vergleichsverfahren ein etwas größeres Spektrum an Therapieverfahren einbezogen wurde, was die Generalisierbarkeit der Ergebnisse erhöht.

9.2.1 Die gegenwartszentrierten Behandlungsverfahren

Eine Gruppe von Therapieverfahren, für die die Evidenzlage noch nicht klar ist, sind die gegenwartszentrierten oder non-direktiven Behandlungsverfahren (PCT). Nach der Hauptanalyse signifikant weniger wirksam als die Vergleichsverfahren, zeigte sich in den Sensitivitätsanalysen ein einflussreicher Fall, dessen Ausschluss wieder äquivalente Ergebnisse bringen würde. Diese Studie (Schnurr et al., 2007) wurde in der vorliegenden Arbeit als bona fide qualifiziert, obwohl Wampold et al. (2010) sie als

nicht-bona fide bezeichnet hatten. Die Begründung für die Aufnahme in den vorliegenden Studienkorpus ist nachvollziehbar und wurde unter 7.1 ausführlich dargelegt. Dies führt aber zu einem Kritikpunkt an der Bona Fide-Definition. Speziell für non-direktive Behandlungsverfahren ist es nach Ansicht der Verfasserin dieser Arbeit keineswegs eindeutig möglich, die notwendigen Kriterien zweifelsfrei zu bestimmen. Zum einen wurde die Bona Fide-Definition auch vom Forscherkreis um Wampold (Benish et al., 2008; Imel et al., 2008) nicht völlig konsistent und überzeugend angewandt, wie am Beispiel zweier Behandlungsverfahren gezeigt wurde (Hedberg & Campbell, 1974; Schnurr et al., 2003). Zum anderen stellt sich die Frage, wie spezifisch ein therapeutisches Verfahren sein muss, um die Kriterien zu erfüllen (siehe 2.3.3).

Eine zusätzliche Information gestaltet die Evidenzlage zu den PCT noch interessanter. Zu Beginn der Analysen war eine Studie im Studienkorpus enthalten (Classen et al., 2011), die wegen fehlender PTSD-Diagnostik, die ein Einschlusskriterium war, wieder ausgeschlossen werden musste. In dieser Studie wurde eine traumafokussierte Gruppentherapie mit einer gegenwartszentrierten Gruppentherapie anhand einer großen Stichprobe ($N = 111$) verglichen. In den um diese Studie erweiterten Analysen zeigten sich die PCT als gleich wirksam wie die Vergleichsverfahren ($g = -.10, p = .256, [KI -0.28, 0.07]$).

Dem folgend könnte man für die PCT zwei Hypothesen aufstellen: Erstens, dass non-direktive Behandlungsverfahren, wenn sie tatsächlich therapeutisch intendiert sind und als gleichwertige aktive Verfahren eingesetzt werden, für die Behandlung der PTSD gleich effektiv sind wie empirisch validierte Therapieverfahren. Diese Hypothese könnte geprüft werden, wenn genügend Vergleichsstudien dieser Art vorliegen. Die zweite Hypothese ist, dass non-direktive Gruppentherapien sich in ihrer Wirksamkeit weniger von anderen Gruppenverfahren unterscheiden als Individualtherapien. Diese Hypothese beruht auf den Ergebnissen zweier großer Studien (Classen et al., 2011; Schnurr et al., 2003) und könnte in Ermangelung genügender Studien im Bereich PTSD störungsbildübergreifend untersucht werden.

Für den gegebenen Studienkorpus ist aber festzuhalten, dass sich die PCT bei der Behandlung der PTSD als weniger wirksam erwiesen, allerdings mit dem Vorteil einer geringeren Dropout-Rate als CBT- oder Expositionsverfahren.

9.2.2 Traumafokussierte versus nicht-traumafokussierte Behandlungsverfahren

Ein weiteres in der aktuellen wissenschaftlichen Diskussion umstrittenes Thema ist die Wirksamkeit traumafokussierter Verfahren im Vergleich zu nicht-traumafokussierten Verfahren. Die Datenlage belegte, dass beide Therapieformen gleich wirksam sind. Die Anzahl der in diesem Vergleich inkludierten Studien war mit 10 Studien und insgesamt 1043 PatientInnen relativ hoch.

Allerdings zeigten sich in den Detail- und Sensitivitätsanalysen auch Trends in Richtung einer größeren Wirksamkeit der traumafokussierten Behandlungsverfahren. Die Gruppe der nicht-traumafokussierten Verfahren bestand zum Großteil aus den gegenwartszentrierten Therapien, wodurch dieses Ergebnis nicht von den oben genannten Ausführungen zu trennen ist, aber auch aus vier anderen Therapieverfahren aus dem kognitiv-verhaltenstherapeutischen Bereich, die keine Fokussierung auf das Trauma beinhalteten.

Das Ergebnis könnte als Beleg dafür betrachtet werden, dass es sich bei diesen Begriffen nicht um eine valide Kategorisierung handelt, die entsprechend operationalisiert ist, wie Wampold et al. (2010) kritisierten. Die andere Erklärung könnte sein, dass die Fokussierung auf das Trauma tatsächlich kein für die Wirksamkeit der Behandlung ausschlaggebender Faktor ist, was wiederum für das Modell der allgemeinen Wirkfaktoren im Sinne von Frank (1997) sprechen würde.

An dieser Stelle sind auch die Analysen zur klinischen Signifikanz zu erwähnen, die ebenfalls einen Trend zu Gunsten der traumafokussierten Therapieverfahren andeuteten. Dieses Ergebnis spricht dafür, unterschiedliche Variablen zur Erfassung von Genesungs- und Besserungsraten differenzierter zu erfassen und auszuwerten (Bradley et al., 2005; Westen & Bradley, 2005), denn es besteht die Möglichkeit, dass in diesen klinischen und für die Praxis bedeutungsvollen Indizes, Unterschiede zwischen den Behandlungsverfahren deutlicher hervortreten.

Nach Auswertung des Studienkorpus der vorliegenden Arbeit kann aber zusammenfassend festgestellt werden, dass auf Basis der momentanen Datenlage kein eindeutiger Unterschied zwischen traumafokussierten und nicht-traumafokussierten Verfahren für direkte Vergleiche zwischen diesen Therapieformen festgestellt werden konnte.

9.2.3 Die Bona Fide-Definition

Trotz aller Kritik an der Bona Fide-Definition sind ihre Vorzüge evident. Sie bietet eine theoretisch fundierte Grundlage, um therapeutisch intendierte Behandlungsverfahren von psychologischen Placebos zu unterscheiden. Gerade für die Frage nach der relativen Effektivität von Psychotherapien ist es essentiell, dass für alle in einer Analyse inkludierten Therapieverfahren die gleichen Ausgangsbedingungen gelten, um einen fairen Vergleich zu ermöglichen. Dass für die Feststellung von empirisch validierten Therapieverfahren der Nachweis von Wirksamkeitsstudien gefordert ist, in denen die Verfahren mit einer Kontrollbedingung verglichen werden, ist ein wesentlicher Grund für die Durchführung dieser Studien. Diese dann allerdings in die Analysen zur relativen Effektivität von Behandlungsverfahren einzubeziehen, führt zu verzerrten Ergebnissen. Dies ist sicher einer der Gründe für die unterschiedlichen Ergebnisse, die BefürworterInnen und KritikerInnen in Bezug auf das Dodo Bird Verdict heranziehen.

Bis jetzt wurde allerdings die Bona Fide-Definition in erster Linie vom Forscherkreis um Wampold angewandt. Eine Ausnahme ist die Studie von Tolin (2010), in der die Kriterien leicht abgewandelt und um ein zusätzliches Kriterium ergänzt wurden. Alle Behandlungsverfahren, in denen die TherapeutInnen in ihren Aktivitäten künstlich eingeschränkt wurden, indem sie beispielsweise vermeiden mussten, über das traumatisierende Ereignis zu sprechen, wurden als nicht-bona fide bewertet. Dieses Kriterium scheint sehr gut geeignet zu sein, um noch deutlicher mittels objektiver Kriterien jene Verfahren zu identifizieren, die als Kontrollbedingung gedacht waren und nicht die Kriterien eines aktiven Therapieverfahrens erfüllen, dem eine Chance auf Erfolg gegeben ist.

Die Bona Fide-Definition könnte auch einem interessanten Forschungsansatz eine operationale Grundlage liefern. Da eine Kritik an der Definition lautet, dass damit impliziert wird, dass Therapieverfahren an sich bona fide sind, ohne zu spezifizieren, für welches Störungsbild sie eingesetzt werden (Ehlers et al., 2010; Siev et al., 2009), könnte eine neue Möglichkeit, das Dodo Bird Verdict zu überprüfen, darin bestehen, dass Therapieverfahren zur Behandlung von Störungsbildern angewandt werden, für die sie eigentlich nicht konzipiert wurden. Ein Beispiel dazu ist der Vergleich von zwei aktiven Therapieverfahren für PatientInnen mit komorbider PTSD

und Alkoholmissbrauchsproblematik, von denen eines der beiden Therapieverfahren nur für eines der beiden Störungsbilder konzipiert wurde (Hien et al., 2004). Ein ähnliches Beispiel ist der Vorschlag, einen Vergleich von dialektisch-behavioraler Therapie, die für Borderline-PatientInnen konzipiert wurde, und *Trauma Model Therapy*, die für die Behandlung von dissoziativen Störungen gedacht ist, für drei PatientInnengruppen mit unterschiedlichen Störungsbildern (Borderline-Persönlichkeitsstörung, Dissoziative Identitätsstörung, beide Störungsbilder) durchzuführen (Ross, 2005). Ein anderer Ansatz dazu könnte sein, Therapieformen wie die Interpersonelle Therapie, die ursprünglich zur Behandlung von Depressionen entwickelt wurde, als Vergleichsverfahren zu nutzen (Schaal, Elbert, & Neuner, 2009).

9.2.4 Das Dodo Bird Verdict

Die Szene aus *Alice im Wunderland*, auf die sich das Dodo Bird Verdict bezieht, handelt von einem Rennen, in dem es keinen gemeinsamen Start und keine Regeln gibt, und das mit dem in der Psychotherapieforschung vielzitierten Spruch des Dodo Vogels beendet wird: „At last the Dodo said, ‘Everybody has won, and *all* must have prizes“ (Carroll, 1920, S. 34). Wampold und seine ForscherkollegInnen (Benish et al., 2008; Imel et al., 2008; Miller et al., 2008; Wampold et al., 1997; Wampold et al., 2002) haben in dieses chaotische Rennen mit der Bona Fide-Definition eine Regel eingeführt und für die MitspielerInnen gleiche Ausgangsbedingungen geschaffen. Ihrer Meinung nach wird jeder Mitspieler, auch wenn er sich noch nie an einem Rennen beteiligt hat, einen Preis gewinnen, wenn er diese Regeln erfüllt. Können die Ergebnisse der vorliegenden Arbeit zu dieser Hypothese einen klärenden Beitrag leisten?

Die Stärke dieser Arbeit liegt darin, dass unter Verwendung der gleichen Datenbasis drei verschiedene metaanalytische Verfahren durchgeführt wurden. Somit kann ein Vergleich dieser Verfahren und ihrer Ergebnisse vorgenommen werden. Der Omnibus-Test wurde für PTSD-Behandlungsverfahren in dieser Arbeit nicht bestätigt. Tatsächlich widerlegen die Ergebnisse dieses Tests das Dodo Bird Verdict. Es ist daher verfrüht, die äquivalenten Ergebnisse, die in den Metaanalysen der Kategorienvergleiche gefunden wurden, auf andere Therapieverfahren zu generalisieren. Für therapeutische Interventionen, die aus kognitiv-behavioralen Modellen entwickelt wurden, ist die Wahrscheinlichkeit, Unterschiede zwischen den Therapien zu finden, gering. Für andere psychotherapeutische Verfahren ist die Datenlage nicht

eindeutig, was auch daran liegt, dass in manchen Kategorien nicht genügend Studien vorhanden sind, um zweifelsfreie Ergebnisse zu erzielen. Insgesamt hat sich keine Kategorie von Behandlungsverfahren als wirksamer erwiesen, nur eine Kategorie als weniger wirksam.

Sind es allgemeine oder spezifische Faktoren, die für diese Ergebnisse verantwortlich sind? Alle inkludierten Therapieverfahren beinhalteten spezifische Techniken. Insofern sind sie für die vorliegende Studienstichprobe tatsächlich ein allgemeiner Wirkfaktor. Dass diese in manchen Therapieverfahren vielleicht nicht in genügendem Ausmaß vorhanden waren, könnte auch als Beleg dafür interpretiert werden, dass spezifische Faktoren tatsächlich einen Unterschied machen. Vielleicht nicht in dem Sinne, welche spezifische Technik verwendet wird, sondern dass sichergestellt werden muss, dass es sich bei der Technik um einen wirkungsvollen Mechanismus handelt (De Rubeis et al., 2005). Das äquivalente Ergebnis des Vergleichs von traumafokussierten und nicht-traumafokussierten Verfahren könnte dafür als Beleg angesehen werden. Um die Medizin-Metapher von Frank (1997) frei zu interpretieren: Am Markt mag es mehrere Präparate geben, die verschiedene Namen tragen, aber alle Aspirin beinhalten. Vielleicht enthält aber eines eine höhere Konzentration des Wirkstoffes und eines einen Wirkstoff schlechterer Qualität. Bei Verwendung als Gerinnungshemmer wäre dies nicht ausschlaggebend, will man aber eine starke entzündungshemmende Wirkung erzielen, dann sehr wohl. Aus diesem Grund schließt sich die Verfasserin dieser Arbeit folgendem Zitat an: „We conclude that the tendency to pit relationship factors against technical ones, or common factors against specific ones, or the dodo bird against ‚empirically supported treatments‘, must be replaced by a more integrative and synergistic perspective“ (Beutler et al., 2004, S. 292).

9.3 Kritik und Grenzen

Welche Kritikpunkte lassen sich für die vorliegende Arbeit finden? Das Einbeziehen der Studie von Schnurr et al. (2007) entgegen der Expertise von Wampold et al. (2010) hatte einen bedeutenden Einfluss auf einzelne Ergebnisse. Die Entscheidung wurde aber ohne genaue Kenntnis der Studienergebnisse wie unter 7.1 beschrieben getroffen.

Die Anzahl der inkludierten Studien in manchen Kategorien war gering, sodass die Teststärke in diesen Kategorien niedrig war und gerade für jene Behandlungsverfahren, die sehr hohe Heterogenitätswerte aufwiesen, keine Moderatorenanalysen möglich waren. Die Moderatoren, die verwendet wurden, haben sich größtenteils als nicht signifikant erwiesen. In diesem Zusammenhang könnte auch ein Unterschied zwischen Moderatoren bestehen, die in Beziehung zu den Ergebnissen einer Therapie gesetzt werden und Moderatoren, die den Unterschied in der Wirksamkeit zweier Behandlungsverfahren aufklären sollen. Insofern wäre es von Vorteil, andere Hypothesen in Bezug auf mögliche Moderatoren zu entwickeln.

Im Gegensatz zu einer Metaanalyse zu alkoholinduzierten Störungen (Imel et al., 2008) wurden für das Random Signs-Verfahren keine Moderatorenanalysen durchgeführt, um die Heterogenität aufzuklären. Das Konzept der Researcher Allegiance wurde ebenfalls nicht als Moderator verwendet, da die Verfasserin dieser Arbeit keine Möglichkeit sah, selbst ein reliables und valides Verfahren zur Erfassung der Einstellung der StudienautorInnen durchzuführen. In der Studie von Tolin (2010) wurde ein neuer Ansatz zur Erfassung der Researcher Allegiance gewählt, in dem die StudienautorInnen direkt kontaktiert wurden. Dies wäre vielleicht auch für diese Studienstichprobe möglich gewesen, da die älteste Studie aus 1989 stammte. Da Researcher Allegiance als eine der wesentlichsten von fünf konfundierenden Variablen in der Erhebung der relativen Effektivität von Therapieverfahren gilt (Staines, 2008), wäre es sicher wünschenswert gewesen, diese Variable in die Moderatorenanalysen einzubeziehen.

Um Moderatoren zu finden, die die fallweise hohe Heterogenität aufklären können, wäre es vielleicht möglich gewesen, TherapeutInnenvariablen, wie beispielsweise eine differenzierte Erfassung der Ausbildung oder des Erfahrungshintergrundes zu erheben, wobei in etwas mehr als der Hälfte aller Studien (12 Studien) die TherapeutInnen beide Behandlungsverfahren durchführten, um die Ergebnisse auf TherapeutInneneffekte zu kontrollieren.

PatientInnenvariablen sind eine weitere konfundierende Größe, die für unterschiedliche Ergebnisse in Wirksamkeitsstudien verantwortlich sein können (Staines, 2008). Motivation, Therapieerwartungen oder die Selbstwirksamkeits-

erwartung wären möglicherweise bedeutende Prädiktoren für den Therapieerfolg, wurden aber schon in den Primärstudien kaum erhoben.

Das Studienqualitätsrating nach Jadad et al. (1996) erwies sich als geeigneter Moderator, um Zusammenhänge zwischen Studienqualität und Effektgrößen sichtbar zu machen. Eine Möglichkeit wäre, noch andere Kriterien zur Untersuchungsmethodik aufzunehmen. "That is, research method factors are a major reason why comparative treatment studies produce discrepant findings about relative efficacy. Such discrepancies can also impede efforts at replication" (Staines, 2008, S. 337).

9.4 Zukünftige Forschung

Mögliche Ansätze für die zukünftige Forschung können in Details liegen oder in Bezug auf die gesamte Frage nach der relativen Effektivität von Psychotherapien richtungsweisend sein.

Im Detail könnte in Metaanalysen mit vergleichbaren Forschungsfragen die Konzeption der untersuchten Variablen noch weiter differenziert werden. Die störungsbezogenen sekundären Maße könnten nach sekundären komorbiden Maßen und sekundären störungsbezogenen Maßen unterschieden werden, und die klinische Signifikanz nach Genesungs- und Besserungsraten, die jeweils für die gesamte Stichprobe und für die Completer-Stichprobe berechnet werden. Wie in dieser Arbeit gezeigt werden konnte, ist die Zusammensetzung der Variablen von entscheidender Bedeutung für das Ergebnis einer Metaanalyse. Insofern wäre es ein Ansatz, verschiedene Variablenkonzeptionen in Bezug auf die Auswirkungen auf den mittleren Effekt zu untersuchen. Da die sekundären Maße durchgängig mehr Heterogenität aufwiesen, könnte durch eine präzisere Variablenkonzeption der sekundären Maße Heterogenität reduziert werden. Einen Beitrag zur Forschung über das Dodo Bird Verdict könnte auch eine Metaanalyse sein, die Therapieverfahren untersucht, die ursprünglich nicht für das behandelte Störungsbild konzipiert wurden (siehe 9.2.3).

Richtungsweisende Vorschläge stammen häufig von AutorInnen, die die Polarität von allgemeinen Wirkfaktoren und spezifischen Techniken überwinden wollen. Mansell (2011) sprach beispielsweise davon, dass es sehr wohl denkbar ist, dass es sowohl allgemeine Wirkfaktoren gibt als auch eine differentielle Wirksamkeit von Psychotherapien. Seiner Meinung nach wäre es notwendig, eine übergeordnete Theorie

über den Prozess der psychologischen Veränderung zu entwickeln, und er verglich dies mit der Theorie der Aerodynamik. Vögel, Fledermäuse, Flugzeuge, andere Flugkörper, alle unterliegen den Gesetzen der Aerodynamik und dennoch gibt es genügend Unterschiede in den verschiedensten Facetten des Fliegens. In seinem Aufruf nach transdiagnostischen und transtheoretischen Zugängen zur Psychotherapieforschung liegt er auf einer Linie mit anderen Forschern, die ihren Fokus zunehmend auf Mediatoren und Mechanismen der Veränderung richten (Kazdin, 2007) oder auf die Erforschung von individualisierten therapeutischen Vorgehensweisen anhand von Entscheidungsbäumen (Longabaugh & Morgenstern, 1999). Insofern wäre es auch aus Sicht der Verfasserin dieser Arbeit eine bereichernde Fokusveränderung, wenn zusätzlich zur Suche nach Unterschieden zwischen Psychotherapien auch die Erforschung der Gemeinsamkeiten von Psychotherapien vertieft würde.

Einer der vier allgemeinen Faktoren, die Rosenzweig 1936 genannt hat (Duncan, 2002), ist der Patient, der im Bereich der Forschung zur relativen Effektivität hauptsächlich in Bezug auf Matching-Hypothesen in Erscheinung tritt. Dies mag vielleicht eine Folge des medizinischen Modells sein, dem Wampold (2001) das kontextabhängige Modell gegenüber stellte. Der Fokus auf Störungsbilder, Diagnosen, Behandlungstheorien und der Entwicklung daraus abgeleiteter spezifischer Techniken führte vielleicht dazu, dass die handelnden Personen, TherapeutInnen und PatientInnen in den Hintergrund gerückt wurden. Insofern ist der Psychotherapieforschung von Wampold (2001) Tribut dafür zu zollen, dieses medizinische Modell kritisch hinterfragt zu haben.

10 Literaturverzeichnis

Mit einem Stern versehene Literaturhinweise sind Studien, die in den Metaanalysen inkludiert sind.

- American Psychiatric Association. (2003). *Diagnostisches und Statistisches Manual Psychischer Störungen - Testrevision (DSM IV-TR, 4th ed.)*. Göttingen: Hogrefe.
- Arindell, W. A., & Ettema, H. (1981). Dimensional structure, reliability and validity of the Dutch version of the Symptom Checklist (SCL-90). *Nederlands Tijdschrift voor de Psychologie*, 36, 77-108.
- Australian Centre for Posttraumatic Mental Health. (2007). *Australian guidelines for the treatment of adults with acute stress disorder and posttraumatic stress disorder*. Melbourne: ACPMH.
- Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, 56, 893-897.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561-571.
- Benish, S. G., Imel, Z. E., & Wampold, B. E. (2008). The relative efficacy of bona fide psychotherapies for treating post-traumatic stress disorder: A meta-analysis of direct comparisons. *Clinical Psychology Review*, 28, 746-758. doi: 10.1016/j.cpr.2007.10.005
- Berg, R. C., & Høie, B. (2010). Effectiveness of psychotherapy for adults with depression: a systematic review of the best available evidence. *Procedia Social and Behavioral Sciences*, 5, 2194-2200. doi: 10.1016/j.sbspro.2010.07.435
- Bernstein, C. E., & Putnam, F. W. (1986). Development, reliability, and validity of a dissociation scale. *Journal of Nervous & Mental Disease*, 174, 727-735.
- Beutler, L. E. (2002). The Dodo bird is extinct. *Clinical Psychology: Science and Practice*, 9 (1), 30-34. doi: 10.1093/clipsy.9.1.30
- Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S., & Wong, E. (2004). Therapist variables. In M. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5th ed., pp. 227-306). New York: John Wiley & Sons.

- Bisson, J., & Andrew, M. (2007). Psychological treatment of post-traumatic stress disorder (PTSD). *Cochrane Database of Systematic Reviews*, 3, Art. No.: CD003388. doi: 10.1002/14651858.CD003388.pub3.
- Bisson, J. I., Ehlers, A., Matthews, R., Pilling S., Richards, D., & Turner, S. (2007). Psychological treatments for chronic post-traumatic stress disorder. *British Journal of Psychiatry*, 190, 97-104. doi: 10.1192/bjp.bp.106.021402
- Bisson, J. I., McFarlane, A. C., & Rose, S. (2000). Psychological Debriefing. In E. B. Foa, T. M. Keane, & M. J. Friedman (Eds.), *Effective treatments of PTSD. Practice guidelines from the International Society for Traumatic Stress Studies* (pp. 39-59). New York: Guilford.
- Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., & Keane, T. M. (1995). The development of a clinician-administered PTSD scale. *Journal of Traumatic Stress*, 8, 75-90. doi: 10.1007/BF02105408
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester, UK: Wiley & Sons Ltd.
- Bortz, J. (1993). *Statistik für Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.
- Bradley, R., Greene, J., Russ, E., Dutra, L., & Westen, D. (2005). A multidimensional meta-analysis of psychotherapy for PTSD. *American Journal of Psychiatry*, 162, 214-227. Retrieved from <http://ajp.psychiatryonline.org/Issue.aspx?issueID=3994&direction=P&journalID=13>
- Braun, S. R. (2012). *Behält der Dodo Recht? Die relative Effektivität von Bona Fide Psychotherapien bei der Behandlung von Depressionen – Verschiedene Metaanalysen* (Master's thesis). In preparation.
- *Brom, D., Kleber, R. J., & Defares, P. B. (1989). Brief psychotherapy for posttraumatic stress disorders. *Journal of Consulting and Clinical Psychology*, 57, 607-612. doi: 10.1037/0022-006X.57.5.607
- *Bryant, R. A., Moulds, M. L., Guthrie, R. M., Dang, S. T., Mastrodomenico, J., Nixon, R. D., . . . Hopwood, S. (2008). A randomized controlled trial of exposure therapy and cognitive restructuring for posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology*, 76, 695-703. doi: 10.1037/a0012616
- Budd, R., & Hughes, I. (2009). The Dodo Bird Verdict – controversial, inevitable and important: A commentary on 30 years of metaanalyses. *Clinical Psychology and Psychotherapy*, 16, 510-522. doi: 10.1002/cpp.648

- Butello, W., & Karl, R. (2011). Dialogic exposure in a gestalt oriented trauma therapy – first RCT results. *European Journal of Psychotraumatology*, 2 (Suppl. 1), 78.
- Carroll, L. (1920). *Alice's adventures in wonderland*. London: Macmillan. Retrieved from http://books.google.at/books?id=Y7sOAAAAIAAJ&printsec=frontcover&dq=inauthor:%22Lewis+Carroll%22&hl=de&ei=OTzdTufQMJP44QTu-tCQDg&sa=X&oi=book_result&ct=result&resnum=8&ved=0CFsQ6AEwBw#v=onepage&q&f=false
- Carroll, K. M., & Rounsaville, B. J. (2010). Perhaps it is the Dodo Bird Verdict that should be extinct. *Addiction*, 105, 14-21. doi: 10.1111/j.1360-0443.2009.02833.x
- Chambless, D. L. (2002). Beware the Dodo bird: The dangers of overgeneralization. *Clinical Psychology: Science and Practice*, 9 (1), 13-16. doi: 10.1093/clipsy/9.1.13
- Chemtob, C. M., Tolin, D. F., van der Kolk, B. A., & Pitman, R. K. (2000). Eye Movement Desensitization and Reprocessing. In E. B. Foa, T. M. Keane, & M. J. Friedman (Eds.), *Effective treatments of PTSD. Practice guidelines from the International Society for Traumatic Stress Studies* (pp. 139-154). New York: Guilford.
- Churchill, R., Hunot, V., Corney, R. Knapp, M., McGuire, H., Tylee, A., & Wessely, S. (2001). A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological treatments for depression. *Health Technology Assessment* 5, 35. Retrieved Sept. 20, 2010 from <http://www.hta.ac.uk/execsumm/summ535.htm>
- Classen, C. C., Palesh, O. G., Cavanaugh, C. E., Koopman, C., Kaupp, J. W., Kraemer, H. C., . . . Spiegel, D. (2011). A comparison of trauma-focused and present-focused group therapy for survivors of childhood sexual abuse: A randomized controlled trial. *Psychological Trauma: Research, Practice, and Policy*, 3, 84-93. doi: 10.1037/a0020096
- Cloitre, M. (2009). Effective psychotherapies for posttraumatic stress disorder: A review and critique. *CNS Spectrums*, 14-1 (Suppl. 1), 32-43. Retrieved from <http://www.cnsspectrums.com/asp/articleDetail.aspx?articleid=1966>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.

- *Cook, J. M., Harb, G. C., Gehrman, P. R., Cary, M. S., Gamble, G. M., Forbes, D., & Ross, R. J. (2010). Imagery Rehearsal for posttraumatic nightmares: a randomized controlled trial. *Journal of Traumatic Stress*, 23, 553-563. doi: 10.1002/jts.20569
- Cook, W. W., & Medley, D. M. (1954). Proposed hostility and pharisaic-virtue scales for the MMPI. *Journal of Applied Psychology* 38, 414-418. doi: 10.1037/h0060667
- *Cottraux, J., Note, I., Yao, S. N., de Mey-Guillard, C., Bonasse, F., Djamoussian, D., . . . Chen, Y. (2008). Randomized controlled comparison of cognitive behavior therapy with rogerian supportive therapy in chronic post-traumatic stress disorder: A 2-year follow-up. *Psychotherapy and Psychosomatics*, 77, 101-110. doi: 10.1159/000112887
- Creamer, M., Bell, R., & Failla, S. (2003). Psychometric properties of the Impact of Event Scale – Revised. *Behaviour Research and Therapy*, 41, 1489-1496. doi:10.1016/j.brat.2003.07.010
- Crits-Christoph, P. (1997). Limitations to the Dodo Bird Verdict and the role of clinical trials in psychotherapy research: Comment on Wampold et al. (1997). *Psychological Bulletin*, 122 (3), 216-220. doi: 10.1037/0033-2909.122.3.216
- Cuijpers, P. (1998). Minimising interventions in the treatment and prevention of depression. Taking the consequences of the 'Dodo Bird Verdict'. *Journal of Mental Health*, 7 (4), 355-365. doi: 10.1080/09638239817950
- 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. doi: 10.1016/j.cpr.2010.06.001
- 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 (6), 909-922. doi: 10.1037/a0013075
- Davidson, P. R., & Parker, K. C. H. (2001). Eye Movement Desensitization and Reprocessing (EMDR): A Meta-Analysis. *Journal of Consulting and Clinical Psychology* 69, (2), 305-316. doi: 10.1037//0022-006X.69.2.305
- Davidson, J. R. T., Smith, R. D., & Kudler, H. S. (1989). Validity and reliability of the DSM-III criteria for posttraumatic stress disorder: Experience with a structured interview. *Journal of Nervous and Mental Disease*, 177, 336-341.

- De Mello, M. F., De Jesus Mari, J., Bacaltchuk, J., Verdeli, H., & Neugebauer, R. (2005). A systematic review of research findings on the efficacy of interpersonal therapy for depressive disorders. *European Archives of Psychiatry and Clinical Neuroscience*, 255, (2), 75-82. doi: 10.1007/s00406-004-0542-x
- DeRubeis, R. J., Brotman, M. A., & Gibbons, C. J. (2005). A conceptual and methodological analysis of the nonspecifics argument. *Clinical Psychology: Science and Practice*, 12 (2), 174-183. doi: 10.1093/clipsy/bpi022
- *Deville, G. J., & Spence, S. H. (1999). The relative efficacy and treatment distress of EMDR and a cognitive-behavior trauma treatment protocol in the amelioration of posttraumatic stress disorder. *Journal of Anxiety Disorders*, 13, 131-157. Retrieved from [http://www.emdr.nl/acrobat/artikelen/Deville%20&%20Spence,%201999%20\(EMDR&TTP\).pdf](http://www.emdr.nl/acrobat/artikelen/Deville%20&%20Spence,%201999%20(EMDR&TTP).pdf)
- Dickersin, K., Chan, S., Chalmers, T. C., Sacks, H. S., & Smith Jr., H. (1987). Publications bias and clinical trials. *Controlled Clinical Trials*, 8 (4), 343-353. doi:10.1016/0197-2456(87)90155-3
- DiGuiseppe, R. (2007). Dissemination of CBT research results: Preaching to the uninterested or engaging in scientific debate. *The Behavior Therapist*, 30 (6), 117-120. Retrieved from <http://www.abct.org/docs/PastIssue/30n6.pdf>
- Dobson, K. S. (1989). A meta-analysis of the efficacy of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, 57 (3), 414-419. doi: 10.1037/0022-006x.57.3.414
- 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 (1), 25-36. doi: 10.1016/j.cpr.2009.08.010
- Duncan, B. L. (2002). The legacy of Saul Rosenzweig: The profundity of the dodo bird. *Journal of Psychotherapy Integration*, 12 (1), 32-57. doi: 10.1037//1053-0479.12.1.32
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455-463. doi: 10.1111/j.0006-341X.2000.00455.x

- Ehlers, A., Bisson, J., Clark, D. M., Creamer, M., Pilling, S., Richards, D., . . . Yule, W. (2010). Do all psychological treatments really work the same in posttraumatic stress disorder? *Clinical Psychology Review, 30*, 269-276. doi: 10.1016/j.cpr.2009.12.001
- Foa, E. B. (1995). *The Posttraumatic Diagnostic Scale (PDS) Manual*. Minneapolis: National Computer Systems.
- *Foa, E. B., Dancu, C. V., Hembree, E. A., Jaycox, L. H., Meadows, E. A., & Street, G. P. (1999). A comparison of exposure therapy, stress inoculation training, and their combination for reducing posttraumatic stress disorder in female assault victims. *Journal of Consulting and Clinical Psychology, 67*, 194-200. doi: 10.1037/0022-006X.67.2.194
- Foa, E. B., Keane, T. M., & Friedman, M. J. (Eds.). (2000). *Effective treatments of PTSD. Practice guidelines from the International Society for Traumatic Stress Studies*. New York: Guilford.
- Foa, E. B., Riggs, D. S., Dancu, C. V., & Rothbaum, B. O. (1993). Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress, 6*, 459-474. doi: 10.1007/BF00974317
- *Foa, E. B., Rothbaum, B. O., Riggs, D. S., & Murdock, T. B. (1991). Treatment of posttraumatic stress disorder in rape victims: A comparison between cognitive-behavioral procedures and counseling. *Journal of Consulting and Clinical Psychology, 59*, 715-723. doi: 10.1037/0022-006X.59.5.715
- Frank, J. D. (1997). *Die Heiler. Wirkungsweisen psychotherapeutischer Beeinflussung. Vom Schamanismus bis zu den modernen Therapien*. Stuttgart: Klett-Cotta.
- Frisch, M. B., Cornell, J., Villanueva, M., & Retzlaff, P. J. (1992). Clinical validation of the Quality of Life Inventory: A measure of life satisfaction for use in treatment planning and outcome assessment. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 4*, 92-101. doi: 10.1037/1040-3590.4.1.92
- 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. doi: 10.1037//0022-006x.63.6.966

- Germain, A., Hall, M., Krakow, B., Shear, M. K., & Buysse, D. J. (2005). A brief sleep scale for posttraumatic stress disorder: Pittsburgh Sleep Quality Index Addendum for PTSD. *Journal of Anxiety Disorders*, 19 (2), 233-244. doi:10.1016/j.janxdis.2004.02.001
- 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 (1), 59-72. doi: 10.1016/s0165-0327(97)00199-7
- Goldberg, D. P., & Hillier, V. F. (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine*, 9 (1), 139-145. doi: 10.1017/S0033291700021644
- Guy, W. (1976). ECDEU Assessment Manual for Psychopharmacology - Revised (pp. 218-222). Rockville: U.S. Department of Health, Education, and Welfare.
- Hamilton, M. (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology*, 32, 50-55. doi: 10.1111/j.2044-8341.1959.tb00467.x
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, 23, 56-62. doi: 10.1136/jnnp.23.1.56
- Hammarberg, M. (1992). Penn Inventory for posttraumatic stress disorder: Psychometric properties. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 4, 67-76. doi: 10.1037/1040-3590.4.1.67
- Harding, T. W., de Arango, M. V., Baltazar, J., Climent, C. E., Ibrahim, H. H., Ladrado-Ignacio, L., . . . Wig, N. N. (1980). Mental disorders in primary health care: a study of their frequency and diagnosis in four developing countries. *Psychological Medicine*, 10, 231-241. doi: 10.1017/S0033291700043993
- Hedberg, A. G. (1974). A comparison of four behavioral treatment approaches to alcoholism. Presentation copy [personal document].
- Hedberg, A. G., & Campbell, L. (1974). A comparison of four behavioral treatments of alcoholism. *Journal of Behavior Therapy and Experimental Psychiatry*, 5 (3-4), 251-256. doi:10.1016/0005-7916(74)90072-X
- Hedges, L. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6, 107-128. doi:10.2307/1164588
- Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

- *Hien, D. A., Cohen, L. R., Miele, G. M., Litt, L. C., & Capstick, C. (2004). Promising treatments of women with comorbid PTSD and substance use disorders. *American Journal of Psychiatry*, 161, 1426-1432. Retrieved from <http://ajp.psychiatryonline.org/Issue.aspx?issueID=3763&direction=P&journalID=13>
- Higgins, J., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539-1558. doi: 10.1002/sim.1186
- Higgins, J., Thompson, S. G., Deeks, J.J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327, 557-560. doi: 10.1136/bmj.327.7414.557
- Horowitz, M. J., Wilner, N. R., & Alvarez, W. (1979). Impact of Event Scale: A measure of subjective stress. *Psychosomatic Medicine*, 41, 209-218. Retrieved from <http://www.psychosomaticmedicine.org/content/41/3/209.full.pdf>
- Howard, K. I., Krause, M. S., Saunders, S. M., & Kopta, S. M. (1997). Trials and tribulations in the meta-analysis of treatment differences: Comment on Wampold et al. (1997). *Psychological Bulletin*, 122, 221-225. Retrieved from http://www.personal.kent.edu/~dfresco/CRM_Readings/Howard_Dodo_1997.pdf
- Huedo-Medina, T., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I² index? *CHIP Documents*, Paper 19. Retrieved from http://digitalcommons.uconn.edu/chip_docs/19
- Hunsley, J., & Di Giulio, G. (2002). Dodo Bird, phoenix, oder urban legend? The question of psychotherapy equivalence. *The Scientific Review of Mental Health Practice* 1 (1). Retrieved April 15, 2010 from <http://www.srmhp.org/0101/psychotherapy-equivalence.html>
- Imel, Z. E., Wampold, B. E., Miller, S. D., & Fleming R. R. (2008). Distinctions without a difference: Direct comparisons of psychotherapies for alcohol use disorders. *Psychology of Addictive Behaviors*, 22 (4), 533-543. doi: 10.1037/a0013171
- *Ironson, G., Freund, B., Strauss, J. L., & Williams, J. (2002). Comparison of two treatments for traumatic stress: A community-based study of EMDR and prolonged exposure. *Journal of Clinical Psychology*, 58, 113-128. Retrieved from <http://www.emdr.nl/acrobat/artikelen/IronsonJCPjanuary.pdf>

- Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J. M., Gavaghan, D. J., & McQuay, H. J. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials*, 17, 1-12. Retrieved from http://www.prosit.de/images/3/36/Assessing_the_Quality_of_Reports_of_Randomized_Clinical_Trials_Is_Blinding_Necessary.pdf
- Janoff-Bulman, R. (1989). Assumptive worlds and the stress of traumatic events: Applications of the schema construct. *Social Cognition*, 7 (2), 113. doi: 10.1521/soco.1989.7.2.113
- Karatzias, A., Power, K., McGoldrick, T., Brown, K., Buchanan, R., Sharp, D., & Swanson, V. (2007). Predicting treatment outcome on three measures for post-traumatic stress disorder. *European Archives of Psychiatry and Clinical Neuroscience*, 257 (1), 40-46. doi: 10.1007/s00406-006-0682-2
- Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *The Annual Review of Clinical Psychology*, 3, 1-27. doi: 10.1146/annurev.clinpsy.3.022806.091432
- Keane, T. M., Caddell, J. M., & Taylor, K. L. (1988). Mississippi Scale for combat-related posttraumatic stress disorder: Three studies in reliability and validity. *Journal of Consulting and Clinical Psychology*, 56, 85-90. doi: 10.1037/0022-006X.56.1.85
- Keane, T. M., Malloy, P. F., & Fairbank, J. A. (1984). Empirical development of an MMPI subscale for the assessment of combat-related posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology*, 52, 888-891. doi: 10.1037/0022-006X.52.5.888
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62, 593-602. Retrieved from http://www.ph.ucla.edu/epi/faculty/detels/PH150/Kessler_DSMIV_AGP2009.pdf
- Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry*, 52 (12), 1048-1060.

- Khawaja, N. G., & Oei, T. P. S. (1992). Development of a catastrophic cognitions questionnaire. *Journal of Anxiety Disorders*, 6, 305-318. doi:10.1016/0887-6185(92)90002-0
- Kilkpatrick, D. G. (1988). Rape aftermath symptom test. In M. Hersen, & A. S. Bellack (Eds.), *Dictionary of behavioral assessment techniques* (pp. 366-367). Oxford, UK: Pergamon Press.
- Krakow, B., Hollifield, M., Schrader, R., Koss, M., Tandberg, D., Lauriello, J., . . . Edmond, T. et al. (2000). A controlled study of imagery rehearsal for chronic nightmares in sexual assault survivors with PTSD: A preliminary report. *Journal of Traumatic Stress* 13 (4), 589-609. DOI: 10.1023/A:1007854015481
- Kubany, E. S., Haynes, S. N., Abueg, F. R., Manke, F. P., Brennan, J. M., & Stahura, C. (1996). Development and validation of the Trauma-Related Guilt Inventory (TRGI). *Psychological Assessment*, 8, 428-444. doi: 10.1037/1040-3590.8.4.428
- Kubinger, K. D. (2009). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens* (2nd ed.). Göttingen: Hogrefe.
- *Lee, C., Gavriel, H., Drummond, P., Richards, J., & Greenwald, R. (2002). Treatment of PTSD: Stress inoculation training with prolonged exposure compared to EMDR. *Journal of Clinical Psychology*, 58, 1071-1089. doi: 10.1002/jclp.10039
- Leichsenring, F. (2001). Comparative effects of short-term psychodynamic psychotherapy and cognitive-behavioral therapy in depression: A meta-analytic approach. *Clinical Psychology Review*, 21 (3), 401-419. doi: 10.1016/s0272-7358(99)00057-4
- Leichsenring, F., & Rabung, S. (2008). Effectiveness of long-term psychodynamic psychotherapy: A meta-analysis. *Journal of the American Medical Association*, 300 (13), 1551-1565. Retrieved from <http://jama.ama-assn.org/content/300/13/1551.full.pdf>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-Analysis*. Applied Social Research Methods Series (Volume 49). Thousand Oaks: Sage Publications.
- Longabaugh, R., & Morgenstern, J. (1999). Cognitive-Behavioral Coping-Skills Therapy for alcohol dependence. Current status and future directions. *Alcohol Research & Health*, 23 (2), 78-85. Retrieved from <http://pubs.niaaa.nih.gov/publications/arh23-2/078-85.pdf>

- Luborsky, L., Rosenthal, R., Diguer, L., Andrusyna, T. P., Berman, J. S., Levitt, J. T., . . . Krause, E. D. (2002). The Dodo Bird Verdict is alive and well – mostly. *Clinical Psychology: Science and Practice*, 9 (1), 2-12. doi: 10.1093/clipsy/9.1.2
- Maercker, A. (2009a). Symptomatik, Klassifikation und Epidemiologie. In A. Maercker (Ed.), *Posttraumatische Belastungsstörungen* (pp. 13-32). Heidelberg: Springer.
- Maercker, A. (2009b). Systematik und Wirksamkeit der Therapiemethoden. In A. Maercker (Ed.), *Posttraumatische Belastungsstörungen* (pp. 137-146). Heidelberg: Springer.
- Mansell, W. (2011). Core processes of psychopathology and recovery: „Does the Dodo bird effect have wings?“ *Clinical Psychology Review*, 31 (2), 189-192. doi: 10.1016/j.cpr.2010.06.009
- *Marks, I., Lovell, K., Noshirvani, H., Livanou, M., & Thrasher, S. (1998). Treatment of posttraumatic stress disorder by exposure and/or cognitive restructuring: A controlled study. *Archives of General Psychiatry*, 55, 317-325. Retrieved from <http://archpsyc.ama-assn.org/content/vol55/issue4/index.dtl>
- Marks, I. M., & Mathews, A. M. (1979). Brief standard self-rating for phobic patients. *Behavior Research and Therapy*, 17, 263-267. doi: 10.1016/0005-7967(79)90041-X
- Marks, I. M., Swinson, R. P., Basoglu, M., Kuch, K., Noshirvani, H., O'Sullivan, G., . . . Sengun, S. (1993). Alprazolam and exposure alone and combined in panic disorder with agoraphobia. A controlled study in London and Toronto. *British Journal for Psychiatry*, 162, 776-787. doi: 10.1192/bjp.162.6.776
- *McDonagh, A., Friedman, M., McHugo, G., Ford, J., Sengupta, A., Mueser, K., . . . Descamps, M. (2005). Randomized trial of cognitive-behavioral therapy for chronic posttraumatic stress disorder in adult female survivors of childhood sexual abuse. *Journal of Consulting and Clinical Psychology*, 73, 515-524. doi: 10.1037/0022-006X.73.3.515
- McLellan, A. T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., Pettinati, H., & Argeriou, M. (1992). The fifth edition of the addiction severity index. *Journal of Substance Abuse Treatment*, 9 (3), 199-213. doi: 10.1016/0740-5472(92)90062-S

- Messer, S. B., & Wampold, B. E. (2002). Let's face facts: common factors are more potent than specific therapy ingredients. *Clinical Psychology: Science and Practice*, 9 (1), 21-25. doi: 10.1093/clipsy.9.1.21
- Miller, S., Wampold, B., & Varhely, K. (2008). Direct comparisons of treatment modalities for youth disorders: A meta-analysis. *Psychotherapy Research*, 18 (1), 5-14. doi: 10.1080/10503300701472131
- Montgomery, S. A., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry*, 134 (4), 382-389. doi: 10.1192/bjp.134.4.382
- Morgenstern, J., & McKay, J. R. (2007). Rethinking the paradigms that inform behavioural treatment research for substance use disorders. *Addiction*, 102, 1377-1389. doi: 10.1111/j.1360-0443.2007.01882.x
- Münker-Kramer, E. (2006). <F43.0> Akute Belastungsreaktion und <F43.1> Posttraumatische Belastungsstörung. In W. Beiglböck, S. Feselmayer, & E. Honemann (Eds.), *Handbuch der klinisch-psychologischen Behandlung* (pp. 293-322). Wien: Springer.
- National Center for Complementary and Alternative Medicine. (2010). Echinacea. [Website] Retrieved from <http://nccam.nih.gov/health/echinacea/>
- National Institute for Clinical Excellence. (2005). Post-traumatic stress disorder (PTSD): The management of PTSD in adults and children in primary and secondary care. Retrieved from <http://www.nice.org.uk/nicemedia/live/10966/29769/29769.pdf>
- *Neuner, F., Schauer, M., Klaschik, C., Karunakara, U., & Elbert, T. (2004). A comparison of narrative exposure therapy, supportive counseling, and psychoeducation for treating posttraumatic stress disorder in an african refugee settlement. *Journal of Consulting and Clinical Psychology*, 72, 579-587. doi: 10.1037/0022-006X.72.4.579
- Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology*, 31 (2), 109-118. doi: 10.1037/h0024436
- *Paunovic, N., & Öst, L. (2001). Cognitive-behavior therapy vs. exposure therapy in the treatment of PTSD in refugees. *Behaviour Research and Therapy*, 39, 1183-1197. doi: 10.1016/S0005-7967(00)00093-0

- Pearlman, L. A. (2001). Treatment of persons with complex PTSD and other trauma-related disruptions of the self. In J. P. Wilson, M. J. Friedman, & J. D. Lindy (Eds.), *Treating psychological trauma and PTSD* (pp. 205-236). New York: Guilford Press.
- Ponniah, K., & Hollon, S. D. (2009). Empirically supported psychological treatments for adult acute stress disorder and posttraumatic stress disorder: A review. *Depression and Anxiety*, 26, 1086-1109. doi: 10.1002/da.20635
- *Power, K., McGoldrick, T., Brown, K., Buchanan, R., Sharp, D., Swanson, V., & Karatzias, A. (2002). A controlled comparison of eye movement desensitization and reprocessing versus exposure plus cognitive restructuring versus waiting list in the treatment of post-traumatic stress disorder. *Clinical Psychology and Psychotherapy*, 9, 299-318. doi: 10.1002/cpp.341
- Powers, M. B., Halpern, J. M., Ferenschak, M. P., Gillihan, S. J., & Foa, E. B. (2010). A meta-analytic review of prolonged exposure for posttraumatic stress disorder. *Clinical Psychology Review*, 30, 635-641. doi: 10.1016/j.cpr.2010.04.007
- Project MATCH Research Group (1997). Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol*, 58, 7-29.
- *Ready, D. J., Gerardi, R. J., Backscheider, A. G., Mascaro, N., & Rothbaum, B. O. (2010). Comparing virtual reality exposure therapy to present-centered therapy with 11 U.S. vietnam veterans with PTSD. *Cyberpsychology, Behavior, and Social Networking*, 13, 49-54. doi: 10.1089/cyber.2009.0239
- Reisner, A. D. (2005). The common factors, empirically validated treatments, and recovery models of therapeutic change. *The Psychological Record*, 55, 377-399. Retrieved from http://findarticles.com/p/articles/mi_hb3538/is_3_55/ai_n29199627/
- *Resick, P. A., Nishith, P., Weaver, T. L., Astin, M. C., & Feuer, C. A. (2002). A comparison of cognitive-processing therapy with prolonged exposure and a waiting condition for the treatment of chronic posttraumatic stress disorder in female rape victims. *Journal of Consulting and Clinical Psychology*, 70, 867-879. doi: 10.1037//0022-006X.70.4.867
- Rogers, C. (1973). *Die klient-bezogene Gesprächstherapie* (2nd ed.). München: Kindler.

- Rosen, C. S., Chow, H. C., Finney, J. F., Greenbaum, M. A., Moos, R. H., Sheikh, J. I., & Yesavage, J. A. (2004). VA practice patterns and practice guidelines for treating posttraumatic stress disorder. *Journal of Traumatic Stress, 17* (3), 213-222. doi: 10.1023/B:JOTS.0000029264.23878.53
- Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry, 6* (3), 412-415. doi: 10.1111/j.1939-0025.1936.tb05248.x
- Ross, C. A. (2005). A proposed trial of dialectical behavior therapy and trauma model therapy. *Psychological Reports, 96*, 901-911. doi: 10.2466/pr0.96.3c.901-911
- *Rothbaum, B. O., Astin, M. C., & Marsteller, F. (2005). Prolonged exposure versus eye movement desensitization and reprocessing (EMDR) for PTSD rape victims. *Journal of Traumatic Stress, 18*, 607-616. doi: 10.1002/jts.20069
- Rothbaum, B. O., Meadows, E. A., Resick, P., & Foy, D. W. (2000). Cognitive Behavioral Therapy. In E. B. Foa, T. M. Keane, & M. J. Friedman (Eds.), *Effective treatments of PTSD. Practice guidelines from the International Society for Traumatic Stress Studies* (pp. 60-83). New York: Guilford.
- Rounsaville, B. J., & Carroll, K. M. (2002). Commentary on Dodo bird revisited: Why aren't we Dodos yet? *Clinical Psychology: Science and Practice, 9* (1), 17-20. doi: 10.1093/clipsy.9.1.17
- Schaal, S., Elbert, T., & Neuner, F. (2009). Narrative Exposure Therapy versus Interpersonal Psychotherapy. *Psychotherapy and Psychosomatics, 78*, 298-306. doi: 10.1159/000229768
- *Schnurr, P. P., Friedman, M. J., Engel, C. C., Foa, E. B., Shea, M. T., Chow, B. K., Resick, P. A., . . . Bernardy, N. C. (2007). Cognitive behavioral therapy for posttraumatic stress disorder in women. A randomized controlled trial. *Journal of the American Medical Association, 297*, 820-830. Retrieved from <http://jama.ama-assn.org/content/297/8.toc>
- *Schnurr, P. P., Friedman, M. J., Foy, D. W., Shea, M. T., Hsieh, F. Y., Lavori, P. W., Glynn, S. M., . . . Bernardy, N. C. (2003). Randomized trial of trauma-focused group therapy for posttraumatic stress disorder. *Archives of General Psychiatry, 60*, 481-489. Retrieved from <http://archpsyc.ama-assn.org/content/vol60/issue5/index.dtl>

- Schnurr, P. P., Friedman, M. J., Lavori, P. W., & Hsieh, F. Y. (2001). Design of Department of Veterans Affairs Cooperative Study No. 420: Group treatment of posttraumatic stress disorder. *Controlled Clinical Trials*, 22, 74-88. doi:10.1016/S0197-2456(00)00118-5
- Seidler, G. H., & Wagner, F. E. (2006). Comparing the efficacy of EMDR and trauma-focused cognitive-behavioral therapy in the treatment of PTSD: a meta-analytic study. *Psychological Medicine*, 36, 1515-1522. doi: 10.1017/S0033291706007963
- Sellman, D. (2010). The 10 most important things known about addiction. *Addiction*, 105 (1), 6-13. doi: 10.1111/j.1360-0443.2009.02673.x
- Shadish, W. R., & Sweeney, R. B. (1991). Mediators and moderators in meta-analysis: There's a reason we don't let dodo birds tell us which psychotherapies should have prizes. *Journal of Consulting and Clinical Psychology*, 59, 883-893. doi: 10.1037//0022-006x.59.6.883
- Shapiro, F. (1995). *EMDR – Grundlagen und Praxis. Handbuch zur Behandlung traumatisierter Menschen*. Paderborn: Junfermann.
- Shapiro, D. A., & Shapiro, D. (1983). Comparative therapy outcome research: Methodological implications of meta-analysis. *Journal of Consulting and Clinical Psychology*, 51 (1), 42-53. doi: 10.1037//0033-2909.92.3.581
- Sheehan, D. V. (1986). *The anxiety disease*. New York: Bantam Books Inc.
- Sherman, J. J. (1998). Effects of psychotherapeutic treatments for PTSD: A meta-analysis of controlled clinical trials. *Journal of Traumatic Stress*, 11 (3), 413-435. doi: 10.1023/A:1024444410595
- Siev, J., & Chambless, D. L. (2007). Specificity of treatment effects: Cognitive therapy and relaxation for generalized anxiety and panic disorders. *Journal of Consulting and Clinical Psychology*, 75 (4), 513-522. doi: 10.1037/0022-006X.75.4.513
- Siev, J., Huppert, J. D., & Chambless, D. L. (2009). The dodo bird, treatment technique, and disseminating empirically supported treatments. *The Behavior Therapist* 32 (4), 69-76. Retrieved from <http://www.abct.org/docs/PastIssue/32n4.pdf>
- Smith, M. L., & Glass G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32 (9), 752-760. Retrieved from http://csrp1.hku.hk/files/975_3792_927.pdf

- Sobell, M. B., Maisto, S. A., Sobell, L. C., Cooper, A. M., Cooper, T., & Sanders, B. (1980). Developing a prototype for evaluating alcohol treatment effectiveness. In L. C. Sobell, M. B. Sobell, & E. Wards (Eds.), *Evaluating alcohol and drug abuse treatment effectiveness: Recent advances* (pp. 129-150). New York: Pergamon.
- Spielberger, C. D. (1988). *State-Trait Anger Expression Inventory: Professional manual*. Odessa: Psychological Assessment Resources.
- Spielberger, C. D., Gorsuch, R. C., & Lushene, R. E. (1970). *Manual for the State Trait Anxiety Inventory*. Palo Alto: Consulting Psychologists Press.
- Staines, G. L. (2008). The relative efficacy of psychotherapy: Reassessing the methods-based paradigm. *Review of General Psychology*, 12 (4), 330-343. doi: 10.1037/1089-2680.12.4.330
- Svartberg, M., & Stiles, T. C. (1991). Comparative effects of short-term psychodynamic psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 59 (5), 704-714. doi: 10.1037/0022-006x.59.5.704
- *Tarrier, N., Pilgrim, H., Sommerfield, C., Faragher, B., Reynolds, M., Graham, E., . . . Barrowclough, C. (1999). A randomized trial of cognitive therapy and imaginal exposure in the treatment of chronic posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology*, 67, 13-18. doi: 10.1037/0022-006X.67.1.13
- Taylor, J. E., & Harvey, S. T. (2009). Effects of psychotherapy with people who have been sexually assaulted: A meta-analysis. *Aggression and Violent Behavior*, 14, 273-285. doi: 10.1016/j.avb.2009.03.006
- *Taylor, S., Thordarson, D. S., Maxfield, L., Fedoroff, I. C., Lovell, K., & Ogradniczuk, J. (2003). Comparative efficacy, speed, and adverse effects of three PTSD treatments: Exposure therapy, EMDR, and relaxation training. *Journal of Consulting and Clinical Psychology*, 71, 330-338. doi: 10.1037/0022-006X.71.2.330
- Tolin (2010). Is cognitive-behavioral therapy more effective than other therapies? A meta-analytic review. *Clinical Psychology Review*, 30, 710-720. doi: 10.1016/j.cpr.2010.05.003
- Van Etten, M. L., & Taylor, S. (1998). Comparative efficacy of treatments for post-traumatic stress disorder: A meta-analysis. *Clinical Psychology and Psychotherapy*, 5, 126-144. doi: 10.1002/(SICI)1099-0879(199809)5:3<126::AID-CPP153>3.0.CO;2-H

- Viechtbauer, W. (2007). Approximate confidence intervals for standardized effect sizes in the two-independent and two-dependent samples design. *Journal of Educational and Behavioral Statistics*, 32 (1), 39-60. doi: 10.3102/1076998606298034
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36 (3), 1-48. Retrieved from <http://www.jstatsoft.org/v36/i03/>
- Wampold, B. E. (2001). *The great psychotherapy debate. Models, methods, and findings*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Wampold, B. E., Imel, Z. E., Laska, K. M., Benish, S., Miller, S. D., Flückinger, C., . . . Budge, S. (2010). Determining what works in the treatment of PTSD. *Clinical Psychology Review*, 30 (8), 923-933. doi: 10.1016/j.cpr.2010.06.005
- 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 – 65. doi: 10.1016/s0165-0327(00)00287-1
- Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, „All must have prizes“. *Psychological Bulletin*, 122 (3), 203-215. doi: 10.1037/0033-2909.122.3.203
- Ware, J. E. Jr., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, 30 (6), 473-483. Retrieved from http://mres.gmu.edu/readings/PSYC557/Ware_Sherbourne_MOS_36item_short-form.pdf
- Watson, C. G., Juba, M. P., Manifold, V., Kucala, T., & Anderson, P. E. (1991). The PTSD interview: rationale, description, reliability, and concurrent validity of a DSM-III-based technique. *Journal of Clinical Psychology*, 47 (2), 179-188. doi: 10.1002/1097-4679(199103)47:2<179::AID-JCLP2270470202>3.0.CO;2-P
- Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993). *The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility*. Paper presented at the Annual Convention of the International Society for Traumatic Stress Studies, San Antonio.

- Weissman, M. M., & Paykel, E. S. (1974). *The depressed woman: A study of social relationships*. Chicago: University of Chicago Press.
- Westen, D., & Bradley, R. (2005). Empirically supported complexity. Rethinking evidence-based practice in psychotherapy. *Current Directions in Psychological Science*, 14 (5), 266-271. doi: <http://dx.doi.org/10.1111/j.0963-7214.2005.00378.x>
- Westmacott, R., & Hunsely, J. (2007). Weighing the evidence for psychotherapy equivalence: Implications for research and practice. *The Behavior Analyst Today*, 8 (2), 128-158. Retrieved from http://findarticles.com/p/articles/mi_6884/is_2_8/ai_n28461012/
- WHO. (2010). *Internationale Klassifikation psychischer Störungen. ICD-10, Kapitel V* (7th ed.). Bern: Huber.
- Willenbring, M. L. (2010). The past and future of research on treatment of alcohol dependence. *Alcohol Research & Health*, 33 (1 and 2), 55-63. Retrieved from <http://pubs.niaaa.nih.gov/publications/arh40/55-63.pdf>
- Wilson, D. B. (2010). Practical Meta-Analysis Effect Size Calculator [Website]. Retrieved from <http://gunston.gmu.edu/cebcp/EffectSizeCalculator/index.html>
- Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, 67 (6), 361-370. doi: 10.1111/j.1600-0447.1983.tb09716.x

11 Tabellenverzeichnis

Tabelle 1	Schematische Einteilung traumatischer Ereignisse nach Maercker, A. (2009a, S. 15)	28
Tabelle 2	Zuordnung der Messinstrumente zur Variable Primäre Maße	43
Tabelle 3	Zuordnung der Messinstrumente zur Variable Sekundäre Maße	44
Tabelle 4	Vergleich von Effektgrößen der primären Maße in Bezug auf den Umgang mit fehlenden Werten	54
Tabelle 5	2x2 Tabelle zur Berechnung des Relativen Risikos	58
Tabelle 6	Erfüllte Qualitätskriterien und Qualitätsbeurteilung der Primärstudien nach Jadad et al. (1996)	73
Tabelle 7	Kategorienzuordnung der inkludierten PTSD-Behandlungsverfahren	77
Tabelle 8	Vergleich von Cohens d auf Basis unterschiedlicher Stichproben	79
Tabelle 9	Ergebnisse des Random Signs-Verfahrens zur relativen Effektivität von Bona Fide-Therapien	83
Tabelle 10	Ergebnisse des Upper Bound-Verfahrens zur relativen Effektivität von Bona Fide-Therapien	86
Tabelle 11	Upper Bound-Verfahren zur relativen Effektivität von Bona Fide-Therapien: Detailanalyse der sekundären Maße	87
Tabelle 12	Kategorienvergleiche zur relativen Effektivität von Bona Fide-Therapien zu Behandlungsende	90
Tabelle 13	Kategorienvergleiche zur relativen Effektivität von Bona Fide- Therapien: Follow-up Daten	92
Tabelle 14	Relatives Risiko zur klinischen Signifikanz zum Behandlungsende	99
Tabelle 15	Relatives Risiko für Dropouts bis zum Behandlungsende	101
Tabelle 16	Moderatorenanalyse: Unterschiede in den Effektgrößen für CBT je nach Studienqualität	103

12 Abbildungsverzeichnis

Abbildung 1	Illustration von John Tenniel (1820 – 1914) aus <i>Alice's adventures in wonderland</i> von Lewis Carroll. .	9
Abbildung 2	Effektgrößen der primären PTSD-Maße aller im Studienkorpus inkludierten Einzelstudien	80
Abbildung 3	Effektgrößen der sekundären Maße aller im Studienkorpus inkludierten Einzelstudien	81

13 ANHANG

- A Abstract / Kurzzusammenfassung
- B Studien in der engeren Wahl, die die Einschlusskriterien nicht erfüllten
(2005 – 2011)
- C Curriculum Vitae

A *Abstract/Kurzzusammenfassung*

Wo sind die Unterschiede? Das Dodo Bird Verdict und die relative Effektivität von Bona Fide-Psychotherapien für die Posttraumatische Belastungsstörung. Eine Metaanalyse.

In einer aktuellen Diskussion steht zur Debatte, ob alle Bona Fide-Psychotherapien für die Posttraumatische Belastungsstörung (PTSD) gleich effektiv sind. Benish, Imel und Wampold (2008) belegten in ihrer Metaanalyse, dass das Dodo Bird Verdict, dass alle Therapien gleich wirksam sind, nicht zurückgewiesen werden kann. Für Ehlers et al. (2010) basierte dieses Ergebnis allerdings auf einer Übergeneralisierung und einer verzerrten Studienauswahl. Andere Metaanalysen fanden äquivalente Ergebnisse für traumafokussierte Behandlungsverfahren, aber nicht für andere Therapien. Zumindest drei Gründe könnten für die unterschiedlichen Ergebnisse verantwortlich sein: 1) die Studienauswahl (und damit die Bona Fide-Definition); 2) die metaanalytischen Verfahren; und 3) die Ergebnismaße. In dieser Diplomarbeit wurde die Benish et al. (2008) Metaanalyse aktualisiert und repliziert und die relative Effektivität von gegenwartszentrierten Therapien untersucht, deren Qualifizierung als Bona Fide-Therapien teilweise unklar ist. Dabei wurden drei metaanalytische Verfahren angewandt, die zeigten, dass bisherige Ergebnisse auf falschen Annahmen beruhten: Das durch eine iterative Prozedur ergänzte Random Signs-Verfahren führte zu einer klaren Zurückweisung des Dodo Bird Verdicts. Die Berechnung eines Upper Bounds, einer Obergrenze für den wahren Unterschied zwischen den Psychotherapien, ähnelte mit $g = 0.23$ den Ergebnissen von Benish et al. (2008). Allerdings handelt es sich dabei, wie empirisch gezeigt werden konnte, eher um ein Maß für die Heterogenität der Therapieeffekte. Metaanalytische Standardvergleiche unterschiedlicher Kategorien von Psychotherapien zeigten, dass sich Expositionsverfahren, kognitive Verhaltenstherapie, EMDR und Prolonged Exposure in ihrer Wirksamkeit nicht von den Vergleichsverfahren unterschieden. Das galt auch für traumafokussierte Therapieverfahren ($g = 0.05$, $p = .542$). Nur die gegenwartszentrierten Therapien waren weniger effektiv ($g = -0.17$, $p = .038$). Aufgrund der geringen Anzahl von Studien in dieser Kategorie sollte dieses Ergebnis jedoch mit Vorsicht interpretiert werden.

Schlüsselbegriffe: Dodo Bird Verdict, Metaanalyse, relative Effektivität, Wirksamkeitsstudien, Posttraumatische Belastungsstörung, Äquivalenzannahme, Psychotherapie

Where are the differences? The dodo bird verdict and the relative efficacy of bona fide psychotherapies in posttraumatic stress disorder: A Meta-analysis.

Currently, there is a debate on whether bona fide psychotherapies differ in efficacy in the treatment of posttraumatic stress disorder (PTSD). Benish, Imel, and Wampold (2008) presented meta-analytical evidence that the Dodo Bird Verdict may not be rejected (i.e., no differences seem to exist). Ehlers et al. (2010) argue that this finding is based on overgeneralization and a biased selection of the available evidence. Other meta-analyses found equivalent efficacy for trauma-focused treatments but not for other therapies. This master's thesis investigated specific issues that seem of importance in this debate. It is argued that differing meta-analytical results may have at least 3 causes: 1) study selection (and thus the bona fide definition); 2) meta-analytical methods; and 3) targeted outcomes. The Benish et al. (2008) meta-analysis was updated and replicated and the relative efficacy of present-centered therapies was also investigated, for which it may seem unclear whether they should be considered bona fide or not. Using three different meta-analytical methods (random signs strategy, computation of an upper bound of true differences between treatments, standard comparisons of different treatment categories) revealed that previous results were likely false. The random signs strategy, augmented with an iterative procedure, led to a clear rejection of the Dodo Bird Verdict. While the computed upper bound, $g = 0.23$, was similar to that obtained by Benish et al. (2008), it is shown empirically that it may only be a measure of the heterogeneity of treatment efficacy. Standard comparisons of different treatment categories revealed that exposure therapy, cognitive-behavioral therapy, EMDR and prolonged exposure did not differ in efficacy to their comparison treatments. This was also true for trauma-focused therapies ($g = 0.05$, $p = .542$). Only present-centered therapy was significantly less effective ($g = -0.17$, $p = .038$). However, due to the only small number of studies in this category this finding should be interpreted with caution.

Key terms: Dodo Bird Verdict, comparative efficacy, relative efficacy, bona fide therapies, PTSD, post-traumatic stress disorder, meta-analysis, treatment studies

B *Studien in der engeren Wahl, die die Einschlusskriterien nicht erfüllten (2005 – 2011)*

Beidel, D. C., Frueh, B. C., Uhde T. W., Wong, N., & Mentrakoski, J. M. (2011). Multicomponent behavioral treatment for chronic combat-related posttraumatic stress disorder: A randomized controlled trial. *Journal of Anxiety Disorders*, 25 (2), 224-231. doi:10.1016/j.janxdis.2010.09.006

⇒ Komponentenstudie

Brett, L., Engel, C. S., Bryant, R. A., & Papa, A. (2007). A randomized controlled proof-of-concept trial of an Internet-based, therapist-assisted self-management treatment for posttraumatic stress disorder. *The American Journal of Psychiatry*, 164 (11), 1676-1683. doi: 10.1176/ appi.ajp.2007.06122057

⇒ nicht-bona fide

Classen, C. C., Palesh, O. G., Cavanaugh, C. E., Koopman, C., Kaupp, J. W., Kraemer, H. C., . . . Spiegel, D. (2011). A comparison of trauma-focused and present-focused group therapy for survivors of childhood sexual abuse: A randomized controlled trial. *Psychological Trauma: Research, Practice, and Policy*, 3, 84-93. doi: 10.1037/ a0020096

⇒ Keine PTSD-Diagnostik

Cloitre, M., Stovall-McClough, K. C., Noonan, K., Zorbas, P., Cherry, S., Jackson, C. L., . . . Petkova, E. (2010). Treatment for PTSD related to childhood abuse: A randomized controlled trial. *The American Journal of Psychiatry*, 167 (8), 915-924. doi: 10.1176/ appi.ajp.2010.09081247

⇒ Komponentenstudie

Gamito, P., Oliveira, J., Rosa, P., Morais, D., Duarte, N., Oliveira, S., & Saraiva, T. (2010). PTSD elderly war veterans: A clinical controlled pilot study. *Cyberpsychology, Behavior, and Social Networking*, 13 (1), 43-48. doi: 10.1089/cyber.2009.0237

⇒ Eine Therapiegruppe umfasste nur zwei Personen

Ginzburg, K., Butler, L. D., Giese-Davis, J., Cavanaugh, C. E., Neri, E., Koopman, C. . . . Spiegel, D. (2009). Shame, guilt, and posttraumatic stress disorder in adult survivors of childhood sexual abuse at risk for human immunodeficiency virus. *The Journal of Nervous and Mental Disease*, 197 (7), 536-542. doi: 10.1097/NMD.0b013e3181ab2ebd

⇒ Zweitartikel (Classen et al., 2011)

Hien, D. A., Wells, E. A., Jiang, H., Suarez-Morales, L., Campbell, A. N. C., Cohen, L. R., . . . Nunes, E. V. (2009). Multisite randomized trial of behavioral interventions for women with co-occurring PTSD and substance use disorders. *Journal of Consulting and Clinical Psychology*, 77 (4), 607-619. doi: 10.1037/a0016227

⇒ nicht-bona fide

Holzapfel, S., Blanchard, E. B., Hickling, E. J., & Malta, L. S. (2005). A crossover evaluation of supportive psychotherapy and cognitive behavioral therapy for chronic PTSD in motor vehicle accident survivors. In M. E. Abelian (Ed.), *Focus on psychotherapy research* (pp. 207-218). Hauppauge, NY: Nova Science Publishers.

⇒ nicht-bona fide

Neuner, F., Onyut, P. L., Ertl, V., Odenwald, M., Schauer, E., & Elbert, T. (2008). Treatment of posttraumatic stress disorder by trained lay counselors in an african refugee settlement: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 76 (4), 686-694. doi: 10.1037/0022-006X.76.4.686

⇒ Laienberater

Nishith, P., Nixon, R. D. V., & Resick, P. A. (2005). Resolution of trauma-related guilt following treatment of PTSD in female rape victims: A result of cognitive processing therapy targeting comorbid depression? *Journal of Affective Disorders*, 86, 259-265. doi: 10.1016/j.jad.2005.02.013

⇒ Zweitartikel zu (Resick et al., 2002)

Roy, M. J., Law, W., Patt, I., Difede, J., Rizzo, A., Graap, K., & Rothbaum, B. (2006). Randomized controlled trial of CBT with virtual reality exposure therapy for PTSD. *Annual Review of CyberTherapy and Telemedicine*, 4, 39-44. Retrieved from <http://www.scribd.com/doc/11648006/Annual-Review-of-CyberTherapy-and-Telemedicine-Volume-4-Summer-2006>

⇒ Komponentenstudie

Schaal, S., Elbert, T., & Neuner, F. (2009). Narrative Exposure Therapy versus Interpersonal Psychotherapy. A pilot randomized controlled trial with Rwandan Genocide Orphans. *Psychotherapy and Psychosomatics*, 78, 298-306. doi: 10.1159/000229768

⇒ Jugendliche

Stapleton, J. A., Taylor, S., & Asmundson, G. J. G. (2006). Effects of three PTSD treatments on anger and guilt: Exposure Therapy, Eye Movement Desensitization and Reprocessing, and Relaxation Training. *Journal of Traumatic Stress*, 19 (1), 19-28. doi: 10.1002/jts.20095

⇒ Zweitartikel zu (Taylor et al., 2003)

Thrasher, S., Power, M., Morant, N., Marks, I., & Dalgleish, T. (2010). Social support moderates outcome in a randomized controlled trial of Exposure Therapy and (or) Cognitive Restructuring for chronic posttraumatic stress disorder. *The Canadian Journal of Psychiatry*, 55 (3), 187-265. Retrieved from <http://publications.cpa-apc.org/media.php?mid=927>

⇒ Zweitartikel zu (Marks et al., 1998)

C *Curriculum Vitae*

Persönliche Daten

Name	Bettina Gregor
Geboren	21.9.1965 in Wien
Staatsbürgerschaft	Österreich
E-Mail	bettina.gregor@inode.at

Schule, Studium und Ausbildungen

1979 - 1984	Höhere Schule für Fremdenverkehrswirtschaft
1994 - 1999	Studium der Psychologie an den Universitäten Innsbruck und Wien (Schwerpunkt Wirtschaftspsychologie)
1994 – 2009	Ausbildung zur Dipl. Lebens- und Sozialberaterin: • NLP: Practitioner & Masterpractitioner bei Trinergy International, Design Human Engineering bei Dr. Richard Bandler, Time Line Therapy bei Tad James • Mediation bei der Arbeitsgemeinschaft für Lösungsorientiertes Konfliktmanagement • Familienaufstellung bei Dr. Brigitte Gross • Systemische Strukturaufstellungen bei Prof. Matthias Varga von Kibed und Dipl.Psych. Insa Sparrer
2004 - 2006	Abgeschlossene Physioenergetik Ausbildung bei Msc Raphael van Assche und Dr. Margot Seitschek Traditionelle Chinesische Medizin bei Claude Diolosa
2008 - 2012	Wiederaufnahme des Studiums der Psychologie an der Universität Wien (Schwerpunkt Klinische und Gesundheitspsychologie)

Berufserfahrung

1986 - 1993	• Novotel Wien West: Abteilungsleitung Rezeption • Vienna Consulting Innsbruck: Unternehmensberatung für Tourismus • Hermes Consulting Innsbruck: Trainingskonzeption
1995/96	Praktikum im Ambulatorium für Schwangerenilfe
1998/99	6-Wochen-Praktikum bei RAT AUF DRAHT
1996 - 2011	Lehraufträge an der Fachhochschule Wiener Neustadt für Kommunikation, Präsentationstechnik, Team-building, Moderations-und Sitzungstechnik, Kreatives Problemlösen, Coaching in der Beratung, Lerntraining
2007 - 2010	Praxis für Physioenergetik im Frauen-Männer-Gesundheitszentrum Trotula
Seit 1993	Selbständig als Trainerin in den Bereichen Tourismus, Persönlichkeitsbildung und Teamentwicklung

Publikationen

Veröffentlichtes Abstract	Gregor, B., & Tran, U. S. (2011). Where are the differences? The dodo bird verdict and the relative efficacy of bona fide psychotherapies in posttraumatic stress disorder: Meta-analysis and comparison with other psychological disorders. <i>European Journal of Psychotraumatology</i> , 2 (Suppl. 1), 144-145.
Posterpräsentation	Gregor, B., & Tran, U. S. (2011). Where are the differences? The dodo bird verdict and the relative efficacy of bona fide psychotherapies in posttraumatic stress disorder: Meta-analysis and comparison with other psychological disorders. <i>12th European Conference on Traumatic Stress, June 2-5 2011, Vienna</i> .

