

**Der Dodo-Mythos in der Psychotherapie -
Sind wirklich alle Behandlungen gleich wirksam?**

**The Dodo Myth in Psychotrauma Therapy -
Are All Treatments Really Equally Effective?**

Oswald J. Klingler

6130 Schwaz

Österreich

+43 6504308045

oswald-j-klingler.at

oswaldklingler@gmail.com

Zusammenfassung

„Jeder hat gewonnen und jeder sollte einen Preis bekommen“, so habe der Riesenvogel Dodo in Alices Wunderland gesagt, und das sollte laut Rosenzweig (1936) auch für die von ihm untersuchten Heilmethoden gelten, welche allesamt als nützlich einzustufen wären. Dieses sogenannte „Dodo-Verdikt“ wurde zum Teil bereitwillig akzeptiert, scheint in der Psychotraumatologie aber in einem deutlichen Gegensatz zu den Ergebnissen der neuesten und umfangreichsten Metaanalysen und den Empfehlungen wichtiger Fachgesellschaften zu stehen, nach denen den sogenannten traumafokussierten gegenüber den gegenwartsfokussierten Behandlungsmethoden der Vorzug zu geben wäre. Den Metaanalysen, von denen dies abgeleitet wird, haftet allerdings der Mangel an, dass in diese auch zahlreiche Vergleichsbedingungen mit methodischen Mängeln und solchen Kontrollbedingungen aufgenommen wurden, welche nicht als vollwertige Therapien eingestuft werden können. Entsprechend wurden in einer eigenen metaanalytischen Untersuchung zur Behandlung von Traumafolgestörungen nur direkte Vergleiche zwischen vollwertigen Behandlungen aufgenommen, deren Durchführung als „bona fide“ also im guten Glauben an deren Wirksamkeit, anzunehmen ist. Dabei fallen im Vergleich der gegenwartsfokussierten mit den traumafokussierten Methoden alle nach der allgemeinen Auswertung als signifikant ($p < 0,01$) einzustufenden Unterschiede zugunsten der traumafokussierten Methoden aus. Bei den gegenwartsfokussierten Methoden lagen mit einer standardisierten Mittelwertsdifferenz von 0,09 im Follow-up bis 0,27 bei Behandlungsabschluss stärkere PTBS-Symptomausprägungen und mit einem relativen Risiko von 1,10 zum Behandlungsabschluss bis 1,64 im Follow-up größere PTBS-Diagnosehäufigkeiten vor. Auch bei einer Einschränkung der Auswertung auf die methodisch am besten zu bewertenden Vergleiche, bei denen kein Verdacht auf eine verzerrende Researcher Allegiance besteht, zeigen alle Signifikanzen Vorzüge der traumafokussierten Verfahren. Der Dodo-Mythos bleibt damit diskussionswürdig.

Schlüsselwörter: Posttraumatische Belastungsstörung, Traumafolgestörungen, Psychotherapie, Meta-Analyse, Forschungsmethoden, Researcher Allegiance.

Abstract

"Everyone has won and everyone should get a prize," said the giant bird Dodo in Alice's Wonderland, and according to Rosenzweig (1936), this should also apply to the healing methods he investigated, all of which could be classified as useful. This so-called "Dodo verdict" was accepted in part, but in psychotraumatology it seems to be in clear contrast to the results of the most recent and comprehensive meta-analyses and the recommendations of important professional societies, according to which so-called trauma-focused treatment methods should be preferred over present-focused treatment methods. The meta-analyses from which this is derived, however, have the shortcoming that they also included numerous treatment comparisons with methodological deficiencies and control conditions that cannot be classified as full-fledged therapies. Accordingly, in a separate meta-analytic study on the treatment of trauma-related disorders, only direct comparisons between full-fledged treatments were included, the implementation of which can be assumed to be "bona fide", i.e. in good faith in their effectiveness. All differences between the present-focused and trauma-focused methods that could be classified as significant ($p < 0.01$) according to the general evaluation were in favor of the trauma-focused methods. Compared to the trauma-focused methods, the present-focused methods had more severe PTSD symptom manifestations, with a standardized mean difference of 0.09 at follow-up to 0.27 at the end of treatment, and a higher frequency of PTSD diagnoses, with a relative risk of 1.10 at the end of treatment to 1.64 at the follow-up. Even if the evaluation is restricted to the comparisons that are methodically best assessed and where there is no suspicion of biasing researcher allegiance, all significances show advantages of the trauma-focused methods. The Dodo myth remains worthy of discussion.

Keywords: Posttraumatic stress disorder, trauma-related disorders, psychotherapy, meta-analysis, research methods, researcher allegiance.

Hintergrund und Fragestellung: Das Dodo-Verdikt und die Traumatherapie

„Jeder hat gewonnen und jeder soll einen Preis bekommen“, so hatte der Riesenvogel Dodo in Lewis Carrolls Alice's Wonderland (1869) gesagt und so sollte das nach dem Psychologen und Psychoanalytiker Saul Rosenzweig (1936) auch für alle Psychotherapieformen gelten. Denn neben der Psychoanalyse könnten auch persuasive Heilmethoden, die Christian Science und zahlreiche andere psychotherapeutische Ideologien auf deutliche Erfolge verweisen.

Es war vermutlich ein Review von Luborsky et al. (1976), der dem so bezeichneten Dodo-Verdikt zu einer breiteren Aufmerksamkeit und Anerkennung in der psychotherapeutischen Fachgesellschaft verholfen hat. Auf Grundlage eines Reviews von insgesamt 112 Behandlungsvergleichen waren die Autoren wegen der eher geringen Anzahl signifikanter Unterschiede zur Schlussfolgerung gekommen, dass man damit Dodos Urteil recht geben müsse. Weiter genährt wurde der Dodo-Mythos, durch Metaanalysen von Wampold (1997) und Luborsky et al. (2002), bei denen allerdings auch manche methodische Probleme auszumachen sind (Klingler, 2023) und dadurch, dass auch in größeren Vergleichsstudien immer wieder keine Unterschiede zwischen den Behandlungseffekten auszumachen sind (z.B. Asarnow und Ougrin, 2017).

Hinsichtlich der Behandlung der posttraumatischen Belastungsstörung (PTBS) schien das Dodo-Verdikt aber bald widerlegt zu sein, zahlreiche Studien und damit auch große Metaanalysen (z. B. nach Watts et al. 2013 und Lewis et al. 2020) zeigten Vorteile zugunsten der sogenannten traumafokussierten Behandlungsmethoden, was auch in entsprechenden Leitlinien (z. B. Schäfer et al. 2019) seinen Niederschlag fand. Doch in die meisten Analysen sind nicht nur Vergleiche zwischen aktiven Behandlungen eingegangen, sondern auch zahlreiche mit Wartegruppen oder „Treatment as Usual“, bei denen auch von wenig wirksamen Behandlungen signifikante Vorteile erwartet werden können (Klingler, 2023). Eine faire Beurteilung von Wirksamkeitsunterschieden würde direkte Vergleiche erfordern, in denen beide Behandlungen „bona fide“, also „im guten Glauben“ an deren Wirksamkeit, durchgeführt wurden. Und bei einer Einschränkung auf solche haben Metaanalysen, wie zum Beispiel von Benish et al. (2008) und Tran und Gregor (2016), nur geringe Wirksamkeitsunterschiede zwischen den Behandlungsmethoden ergeben. Tran und Gregor

konnten in ihre Analyse 24 Vergleiche von Behandlungen einbeziehen, bei denen die Durchführung von beiden als bona fide beurteilt wurde. In einem zusammenfassenden Vergleich von so bezeichneten traumafokussierten (TFT) mit gegenwartsfokussierten (GFT) Therapien wurde dann wohl eine statistisch signifikante Überlegenheit für die TFT ermittelt, allerdings mit relativ geringen Effektstärken von 0,14, beziehungsweise von 0,19 und 0,23 im Follow-up.

Ob damit das Dodo-Verdikt bestätigt oder widerlegt wird, wird von Tran und Gregor (2016) offen gelassen. Sie verweisen auf eine Reihe von methodischen Problemen und - der Komplexität des Gegenstandes entsprechend - ihre Arbeit wirft auch eine Reihe von Fragen auf. So könnte eine Unterschätzung von Wirksamkeitsunterschieden erfolgt sein, dadurch, dass in die TFT-GFT-Vergleiche nicht nur Behandlungen im Einzelsetting, sondern auch solche im Gruppensetting einbezogen wurden, welche aber für die TFT-Methoden allgemein weniger geeignet sein dürften (vergl. z.B. Watts et al. 2013). Auch scheint der Einfluss der Methodik und der sogenannten „Researcher Allegiance“ noch einer eingehenderen Untersuchung bedürftig. Nach Tran und Gregor haben sich zwar bei einer als günstiger zu beurteilenden methodischen Qualität der Studien geringere Effektstärken gezeigt, werden allerdings ihre Vergleiche auf die methodisch höchstwertigsten eingeschränkt werden, dann bleiben nur mehr die Arbeit von Schnurr et al. (2007), mit einem sehr deutlichen positiven Effekt einer traumafokussierten Einzelbehandlung, und jene von Cook et al. (2010), in der sich im Gruppensetting keine Unterschiede eines „Imagery Rehearsal“ und einer verhaltenstherapeutischen Vergleichsbehandlung gezeigt haben. Und obwohl sie auf die Ergebnisse von Munder et al. (2012) hingewiesen haben, die einen Einfluss der sogenannten „Researcher Allegiance“, also der Verbundenheit der Forscherinnen und Forscher mit einer der geprüften Behandlungen, auf die Ergebnisse von Vergleichsstudien gezeigt haben, wurde dieser von Tran und Gregor nicht in Hinblick auf die von ihnen verwendeten Studien geprüft. Für die Praxis der Traumatherapie dürfte der Vergleich der traumafokussierten mit den gegenwartsfokussierten Methoden von besonderer Bedeutung sein. Werden doch die TFT von den bedeutendsten Fachgesellschaften empfohlen, aber von nicht wenigen Therapeutinnen und Therapeuten vermieden, wegen Bedenken hinsichtlich möglicher Retraumatisierungen und Behandlungsabbrüche. Die Frage, ob es hier Vorteile für bestimmte

Methoden gibt oder ob doch wieder das sogenannte Dodo-Verdikt - alle bekommen einen Preis - akzeptiert werden muss, hat hier also durchaus Berechtigung.

Methoden

Es wurde eine Pub-Med-Recherche mit den Stichwörtern „traumatic“, „stress“, „disorder“, „psychotherapy“ und dem Kriterium „randomized trial“ durchgeführt, welche mit Stichtag vom 02.01.2025 insgesamt 1823 Treffer mit bis Ende 2024 dokumentierten englischsprachigen Publikationen ergeben (keine auf Deutsch). Davon wurden sukzessive (ohne Prüfung etwaiger zusätzlicher Ausschließungsgründe) insgesamt 1715 Arbeiten ausgeschlossen, aus den in Tabelle 1 ausgewiesenen Gründen.

Ausschließungsgründe:	Anzahl ausgeschlossener Treffer:
Mehrfachpublikationen, Mehrfachnennungen (E-Publishing)	220
keine randomisierte Vergleichsstudie	588
sekundär durchgeführte Subgruppenvergleiche	17
keine Durchführung der verglichenen Behandlungen im guten Glauben (bona fide) an deren Wirksamkeit anzunehmen (Ausschluss von Vergleichen mit Wartegruppen, „TAU“, mit Behandlungen, die mit deutlich geringerem Aufwand betrieben wurden, mit Behandlungen, die als „Kontrollen“ deklariert wurden)	583
kein Vergleich mit einer psychotherapeutischen Behandlung	92
Vergleich von unterschiedlichen Behandlungssettings	50
keine Trauma-Folgestörung behandelt, beziehungsweise kein Trauma oder keine Traumasymptome gemäß DSM oder ICD dokumentiert	107
keine der für eine metaanalytische Auswertungen erforderlichen Daten vorhanden (vergleiche Abschnitt 3.2.4)	58
Tabelle 1: Von der weiteren Auswertung ausgeschlossene Treffer der Recherche.	

Die nach der PubMed-Recherche so verbleibenden 108 Arbeiten wurden schließlich um 15 zusätzliche erweitert, auf welche nach den Literaturhinweisen in den PubMed-Treffern und nach den Recherchen von Lewis et al. (2020) und Klingler (2023a) zurückgegriffen werden konnte. Das ergab in Summe 123 einzubeziehende Studien mit insgesamt 149 Vergleichen

von Behandlungen, deren Durchführung als bona fide, also im guten Glauben an deren Wirksamkeit angenommen werden kann. Unter diesen waren 63 Vergleiche von gegenwartsfokussierten (GFT) mit traumafokussierten Behandlungsmethoden (TFT), 45 davon mit Behandlungen eines Vollbildes einer PTBS, davon 42 bei Erwachsenen und 41 im Einzelsetting. Einen Überblick über die einbezogenen Vergleiche mit Literaturhinweisen bietet der Anhang. Für die Aufnahme als bona fide Behandlungen sind dabei striktere Kriterien als bei Tran und Gregor (2016) zur Anwendung gekommen; so wurden beispielsweise die schon erwähnten Arbeiten von Schnurr et al. (2007) und Cook et al. (2007) nicht einbezogen, weil bei diesen jeweils eine der verglichenen Behandlungen als „control“ bezeichnet wurde. Eine Beschreibung der einbezogenen Studien findet sich im Anhang (oswald-j-klingler.at).

Bei den einbezogenen Vergleichen sollte die Beurteilung der Behandlungseffekte, so weit wie vorhanden, auf Grundlage der folgenden Variablen erfolgen:

- Anzahl der Behandlungsabbrüche („Abbrecher“),
 - Anzahl der Personen, bei welchen nach der Behandlung die Kriterien einer PTBS-Diagnose erfüllt waren („PTBS-Diagnosen“),
 - Ausprägung von PTBS-Symptomen, Selbstbeurteilungsverfahren,
 - Ausprägung von PTBS-Symptomen, Fremdbeurteilungsverfahren,
 - Ausprägung von depressiven Symptomen, Selbstbeurteilungsverfahren,
 - Ausprägung von depressiven Symptomen, Fremdbeurteilungsverfahren,
- und außer den Abbrüchen alle Variablen in ihren Ausprägungen
- hinsichtlich des Messzeitpunktes unmittelbar nach Behandlung („post“) oder „Follow-up“ („FU“)
 - hinsichtlich der ausgewerteten Personen, ob alle, welche die Behandlung begonnen haben („intent to treat“: „ITT“) oder alle, welche die begonnene Behandlung auch abgeschlossen haben („observed cases“: „OC“).

Damit waren Auswertungen auf insgesamt $1+4+4+4+4+4 = 21$ Variablen vorgesehen, nämlich für die Anzahl der Abbrecher und für die vorhandenen PTBS-Diagnosen und die Ausprägungen PTBS-Symptome/Selbstbeurteilung, PTBS-Symptome/Fremdbeurteilung, depressive Symptome/Selbstbeurteilung und depressive Symptome/Fremdbeurteilung

jeweils in den Varianten post/ITT, post/OC, FU/ITT, FU/OC. Die Prüfung von so vielen Variablen scheint erforderlich, weil nicht auszuschließen ist, dass manche Einflüsse eher die Selbst- und andere eher die Fremdbeurteilungsverfahren betreffen, manche Behandlungen eher kurz- und andere eher längerfristig wirken und manche eher bei jenen, welche die Behandlung auch abschließen („observed cases“). Mit einer Zusammenfassung beziehungsweise Vermischung dieser Variablen in der Auswertung bestünde aber die Gefahr eines Verrauschens beziehungsweise einer Unterschätzung von Therapieeffekten.

Für jeden Behandlungsvergleich wurden Qualitätsscores (QS) hinsichtlich der folgenden Kriterien gebildet:

- QS1: angemessener Vergleich - Behandlungsaufwand und Behandlungsdauer bei den verglichenen Behandlungen gleichwertig;
- QS2: vergleichbare Behandlungsqualität - Ausbildung der Therapeutinnen/Therapeuten und geprüfte Behandlungsqualität vergleichbar;
- QS3: ausreichende Behandlerunabhängigkeit - mehr als eine Therapeutin/ein Therapeut pro Behandlungsmethode;
- QS4: geprüfte Vergleichbarkeit der Behandelten - Ausgangswerte und mindestens 7 prognostische Variablen ohne Unterschiede mit $p < 0,05$;
- QS5: keine Erwartungs-Differenzen: „treatment credibility“ aus Sicht der Behandelten geprüft (z. B. in Anlehnung an Borkovec & Nau 1972) und vergleichbar;
- QS6: Dokumentation von Abbrechern;
- QS7: „Blindbeurteilung“ der abhängigen Variablen/Ergebnisse.

Von einer Researcher-Allegiance zugunsten einer der verglichenen Behandlungen wurde ausgegangen, wenn unter den Autorinnen oder Autoren der Arbeit nur solche auszunehmen waren, von denen durch Publikation von entsprechenden Behandlungskonzepten oder Manualen (nicht von Effektivitätsstudien) eine besondere Verbundenheit mit einer dieser Behandlungen angenommen werden kann, aber keine andere Person, mit einer entsprechenden Verbundenheit mit der Vergleichsbehandlung.

Die Analysen wurden mithilfe des Programmpakets OpenMetaAnalyst (<http://www.cebm.brown.edu/openmeta/>) unter Anwendung eines Random-Effects-Modells nach Hedges-Olkin durchgeführt, mit den relativen Risiken (RR) für die diskreten Variablen

„Abbrecher“ und „PTBS-Diagnosen“ und den standardisierten Mittelwertsdifferenzen (SMD) für die stetigen Variablen „PTBS-Symptome“ und „depressive Symptome“ als Effektgrößen in den schon genannten Varianten „ITT“ und „OC“, „post“ und „FU“, für jede Variable über jene Vergleiche, aus welchen hierfür die Daten verfügbar waren. Zunächst erfolgte für alle abhängigen Variablen eine regressionsanalytische Prüfungen des Einflusses der Qualitätskriterien Q1 bis Q7 und der Allegiance. Dann wurde untersucht, welche Unterschiede zwischen GFT und TFT allgemein bestehen. Weiters wurden Subgruppenanalysen durchgeführt mit Subgruppen, welche aus den Diagnosen (vollständige oder unvollständige PTBS), Setting (Einzel oder Gruppe), Alter (Adult oder Kinder/Jugendliche) gebildet wurden, ergänzt durch solche über die Stufen der Methodenkriterien und der Allegiance, wenn sich für diese regressionsanalytisch signifikante Einflüsse ergeben haben. Schließlich wurde auch noch geprüft, welche Unterschiede bleiben, wenn in einem „Modell der besten Vergleiche“ lediglich jene Vergleiche berücksichtigt werden, welche methodisch und hinsichtlich der Researcher-Allegiance am günstigsten zu beurteilen sind. Hierfür wurden aus den Vergleichen mit neutraler Allegiance jene mit der besten methodischen Bewertung ausgewählt, wobei die geforderten QS-Scores so festgelegt wurden, dass Daten aus zumindest drei Vergleichen verwendet werden konnten. Das erforderliche Signifikanzniveau wurde für alle Auswertungen mit $p < 0,01$ gesetzt. Weitere Details der Methodik können dem Anhang (oswald-j-klingler.at) entnommen werden.

Ergebnisse

Als signifikante Einflüsse der Methodenkriterien liegen solche von QS4 (geprüfte Vergleichbarkeit von prognostischen Variablen) auf die Häufigkeit von PTBS-Diagnosen (FU/ITT und FU/OC) und von QS7 (Blindbeurteilung der abhängigen Variablen) auf die Ausprägung von PTBS-Symptomen (FU/ITT) vor. Mit diesen bestehen bei einer positiven Beurteilung der Methodik jeweils günstigere Werte für die TFT. Ein signifikanter Einfluss der Researcher Allegiance ist hinsichtlich der Ausprägung von PTBS-Symptomen (Fremdbeurteilung/FU/ITT) und depressiven Symptomen (Selbstbeurteilung/post/ITT und

OC) gegeben, jeweils mit günstigeren Werten für jene Behandlung, die als bevorzugt gelten kann.

Alle GFT-TFT-Vergleiche mit signifikanten Unterschieden werden in Tabelle 1 zusammengefasst.

Gesamtheit/ Subgruppe:	Abhängige Variable:	Vergl.:	N:	ES:	Vorteil:
Total	PTBS-Diagnosen/post/OC	21	1035	RR 1,37!	TFT
	PTBS-Diagnosen/FU/OC	15	716	RR 1,64!!	TFT
	PTBS-Symptome/Fremdbeurteilung/post/ITT	15	1170	SMD +0,27!	TFT
	PTBS-Symptome/Fremdbeurteilung/post/OC	33	1950	SMD +0,15!	TFT
	PTBS-Symptome/Selbstbeurteilung/FU/ITT	9	645	SMD +0,23!	TFT
	PTBS-Symptome/Fremdbeurteilung/FU/OC	27	1443	SMD +0,23!	TFT
Modell der besten Vergleiche	PTBS-Diagnosen/FU/ITT	5	589	RR 1,27!	TFT
	PTBS-Diagnosen/FU/OC	3	239	RR 1,88!	TFT
	PTBS-Symptome/Fremdbeurteilung/FU/OC	7	452	SMD +0,33!	TFT
Vollbild PTBS	PTBS-Diagnosen/post/OC	17	813	RR 1,45!	TFT
	PTBS-Diagnosen/FU/OC	11	497	RR 1,61!	TFT
	PTBS-Symptome/Fremdbeurteilung/FU/OC	22	987	SMD +0,31!!	TFT
Symptome/ unvollständige PTBS	PTBS-Symptome/Fremdbeurteilung/post/ITT	4	347	SMD +0,60!	TFT
	PTBS-Symptome/Selbstbeurteilung/FU/ITT	3	178	SMD +0,52!	TFT
Einzelsetting	PTBS-Diagnosen/post/OC	21	1035	RR 1,37!	TFT
	PTBS-Diagnosen/FU/OC	15	716	RR 1,64!!	TFT
	PTBS-Symptome/Fremdbeurteilung/post/ITT	33	1950	SMD +0,15!	TFT
	PTBS-Symptome/Fremdbeurteilung/FU/OC	27	1443	SMD +0,23!	TFT
Gruppensetting	Abbrecher	2	268	RR 0,50!!	GFT
Eltern-Kind-Setting	PTBS-Diagnosen/post/ITT	2	257	RR 1,47!	TFT
Erwachsene	PTBS-Diagnosen/post/OC	20	1006	RR 1,43!	TFT
	PTBS-Diagnosen/FU/OC	15	716	RR 1,64!!	TFT
	PTBS-Symptome/Fremdbeurteilung/post/ITT	12	1042	SMD +0,32!	TFT
	PTBS-Symptome/Fremdbeurteilung/post/OC	31	1855	SMD +0,16!	TFT
	PTBS-Symptome/Fremdbeurteilung/FU/OC	26	1379	SMD +0,25!!	TFT
QS4: prognostische Variablen vergleichbar	PTBS-Diagnosen/FU/ITT	14	854	RR 1,37!	TFT
	PTBS-Diagnosen/FU/OC	11	396	RR 1,84!	TFT
QS7: Blindbeurteilung	PTBS-Symptome/Selbstbeurteilung/FU/ITT	8	530	SMD +0,33!!	TFT
Allegiance pro TFT	PTBS-Diagnosen/post/ITT	8	541	RR 1,30!!	TFT
	Depress. Sympt./Selbstbeurteilung/post/ITT	4	223	SMD +0,56!!!	TFT

Tabelle 1: Signifikante Ergebnisse der GFT-TFT-Vergleiche, die RR>1 beschreiben größere Häufigkeiten, SMD>0 stärkere Symptomausprägungen unter GFT (p<0,01: !, p<0,001: !!, p<0,0001: !!!).

Innerhalb der Gesamtheit der verfügbaren Vergleiche („Total“) fallen alle signifikante Unterschiede zugunsten der TFT aus. Die einzige Subgruppe, in der bessere Werte für die GFT festzustellen sind, ist das Gruppensetting, in dem nach 2 Studien eine geringere Zahl an

Behandlungsabbrüchen als bei TFT vorliegt. Das hier gegebene geringere Risiko unter GFT muss vornehmlich auf eine erhöhte Anzahl von Abbrüchen unter TFT im Gruppensetting zurückgeführt werden (bei TFT im Gruppensetting 30 Prozent und im Einzelsetting 21 Prozent Behandlungsabbrüche).

Hinsichtlich der Größenordnung der Effektstärken bestehen recht deutliche Unterschiede nach den jeweiligen abhängigen Variablen. Sie liegen in der allgemeinen Auswertung zwischen RR 1,10 und 1,64 für die Häufigkeit der PTBS-Diagnosen, zwischen SMD +0,04 und +0,27 für die Ausprägung der PTBS-Symptome und SMD -0,13 und +0,34 für die depressiven Symptome.

Nicht so deutlich wie erwartet zeigte sich der Einfluss der Methodenkriterien. Dieser ist auch nicht einfach interpretierbar. Dass sich bei einer geprüften Vergleichbarkeit von Ausgangswerten und prognostischen Variablen (QS4) und bei einer Blindbeurteilung von abhängigen Variablen (QS7) günstigere Werte für die TFT ergeben, scheint aus den Daten kaum erklärbar. Auf eine Prüfung der Möglichkeit, dass bei einer Allegiance für traumafokussierte Behandlungsmethoden eine größere Sorgfalt bezüglich Forschungsmethoden vorliegen könnte, wurde wegen der geringen Anzahl entsprechender Vergleiche verzichtet. Dass von methodisch sorgfältiger geplanten und durchgeführten Arbeiten verlässlichere und gültigere Ergebnisse zu erwarten wären, darf aber auch ohne eindeutige Belege durch die Daten angenommen werden.

Ein anschauliches Beispiel für den Einfluss der Researcher Allegiance kann mit dem metaanalytischen Ergebnisplot in Abbildung 1 geboten werden. Dieser zeigt günstigere Werte in den Subgruppen mit entsprechender Allegiance. Bei Ausschluss von Arbeiten mit erkennbarer Researcher-Allegiance und geringerer methodischer Qualität im Rahmen einer Auswertung der besten Vergleiche liegen eher etwas höhere Effektstärken vor, nämlich zwischen RR 1,15 und 2,00 für die diskreten und SMD +0,03 und +0,35 für die stetigen Variablen.

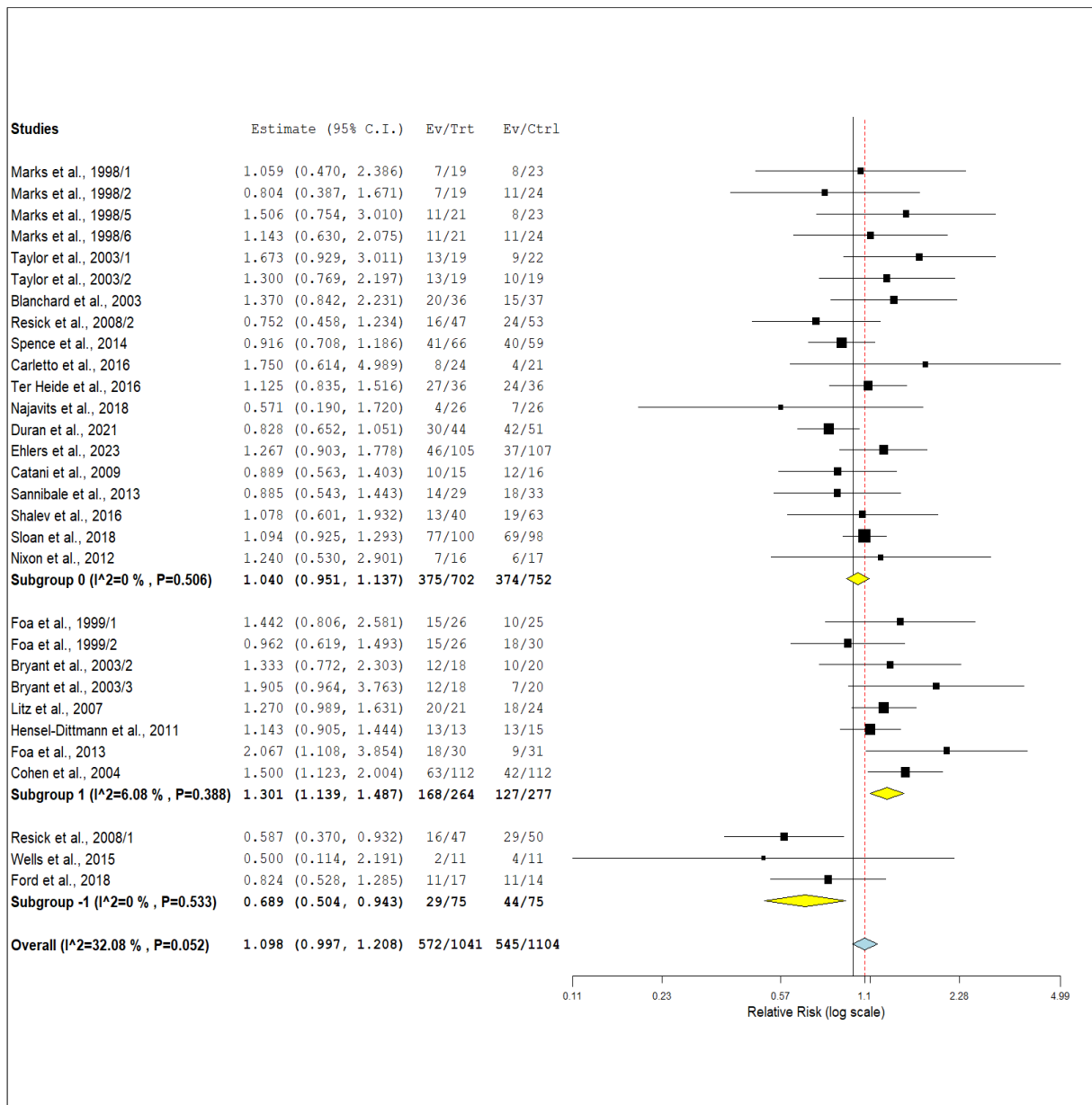


Abbildung 1: Relative Risiken GFT versus TFT für PTBS-Diagnosen/post/ITT nach Researcher-Allegiance (Subgroup 0: Allegiance neutral, Subgroup 1: Allegiance für TFT, Subgroup -1: Allegiance für GFT; geringere relative Risiken unter 1 beschreiben eine geringere Diagnosehäufigkeit unter GFT). Beschreibung der einbezogenen Studien im Anhang (oswald-j-klingler.at).

Den Abbrecherzahlen sei noch ein wenig zusätzliche Aufmerksamkeit gewidmet, wird doch ein Behandlungsabbruch für Betroffene fast immer eine Belastung und ein Risiko darstellen. Im Gruppensetting ist eine hohe Abbrecherzahl bei traumafokussierter Behandlung keine Überraschung. Doch auch im Einzelsetting sind unter TFT die Abbrecherzahlen etwas erhöht, wenn auch in einem nicht signifikanten Ausmaß (vergleiche Abbildung 2). Für einen Vergleich von Abbrecherzahlen und PTBS-Häufigkeiten wurden nun die methodisch bestbewerten

Studien ausgewählt, bei denen im Einzelsetting entsprechende Werte vorliegen. Das waren die Arbeiten von Resick et al. (2002), Taylor et al. (2003) und Ehlers et al. (2023), wobei bei Taylor et al. die Daten aus zwei Vergleichen und bei Resick et al. nur ITT-Daten verwendet werden konnten.

Bei den 4 Vergleichen mit vorliegender ITT-Auswertung stehen geringfügig erhöhten Abbrecherzahlen unter TFT nur geringfügig erhöhten PTBS-Häufigkeiten unter GFT gegenüber. Sehr viel deutlicher sind die Unterschiede bei den besten Vergleichen mit OC-Auswertung. Hier liegen in der Zusammenfassung der beiden Vergleiche von Taylor et al., 2003 (eine Gruppe GFT, 2 Gruppen TFT), und dem Vergleich von Ehlers et al., 2023, nach der Behandlung sehr viel mehr PTBS-Diagnosen unter GFT vor: nämlich bei 124 Randomisierten und 13 Abbrechern 38 zum Zeitpunkt post und 32 zum Zeitpunkt FU, unter TFT hingegen bei 148 Randomisierten und 20 Abbrüchen nur 23 zum Zeitpunkt post beziehungsweise 21 zum Zeitpunkt FU. Das entspricht signifikanten relativen Risiken von 2,09 für den Zeitpunkt post und 1,94 für den Zeitpunkt FU (prozentuale Darstellung in den Abbildungen 3a für ITT-Daten und 3b für die OC-Daten). Wird eine traumafokussierte Behandlung also vollständig absolviert, dann ist die Chance eines „loss of diagnosis“ etwa doppelt so groß wie bei einer gegenwartsfokussierten Behandlung.

Eine detailliertere Beschreibung der Ergebnisse kann dem Anhang (oswald-j-klingler.at) entnommen werden.

Studies	Estimate (95% C.I.)	Ev/Trt	Ev/Ctrl
Brom et al., 1989	1.06897 (0.29423, 3.88370)	4/29	4/31
Carlson et al., 1998	2.35714 (0.10601, 52.41052)	1/13	0/10
Marks et al., 1998/1	1.21053 (0.27543, 5.32024)	3/19	3/23
Marks et al., 1998/2	0.94737 (0.24058, 3.73055)	3/19	4/24
Marks et al., 1998/5	0.15584 (0.00852, 2.84985)	0/21	3/23
Marks et al., 1998/6	0.12626 (0.00719, 2.21582)	0/21	4/24
Foa et al., 1999/1	3.36538 (0.77190, 14.67272)	7/26	2/25
Foa et al., 1999/2	1.00962 (0.42389, 2.40469)	7/26	8/30
Bryant et al., 2003/2	0.66667 (0.18500, 2.40234)	3/18	5/20
Bryant et al., 2003/3	0.66667 (0.18500, 2.40234)	3/18	5/20
Taylor et al., 2003/1	0.66165 (0.22829, 1.91766)	4/19	7/22
Taylor et al., 2003/2	1.00000 (0.29188, 3.42605)	4/19	4/19
Blanchard et al., 2003	0.92500 (0.42620, 2.00756)	9/36	10/37
Coffey et al., 2006	0.33333 (0.01491, 7.45063)	0/12	1/12
Litz et al., 2007	0.45714 (0.16805, 1.24359)	4/21	10/24
Cottraux et al., 2008	4.64286 (1.47765, 14.58809)	13/28	3/30
Resick et al., 2008/1	0.88652 (0.42352, 1.85572)	10/47	12/50
Resick et al., 2008/2	1.02515 (0.47894, 2.19429)	10/47	11/53
Ready et al., 2010	1.20000 (0.09802, 14.69039)	1/5	1/6
Hensel-Dittmann et al., 2011	0.76923 (0.15105, 3.91730)	2/13	3/15
Spence et al., 2014	0.07449 (0.00999, 0.55573)	1/66	12/59
Markowitz et al., 2015	0.42105 (0.14213, 1.24733)	4/38	9/36
Rauch et al., 2015	0.42857 (0.13113, 1.40067)	3/18	7/18
Wells et al., 2015	1.00000 (0.07116, 14.05270)	1/11	1/11
Carletto et al., 2016	1.75000 (0.17062, 17.94898)	2/24	1/21
Ter Heide et al., 2016	1.33333 (0.51443, 3.45580)	8/36	6/36
Ford et al., 2018	0.54902 (0.25861, 1.16554)	6/17	9/14
Najavits et al., 2018	0.57143 (0.18990, 1.71951)	4/26	7/26
Nidich et al., 2018	1.36364 (0.67603, 2.75061)	15/68	11/68
Norman et al., 2019	0.50000 (0.28082, 0.89026)	12/56	27/63
Thorp et al., 2019	0.78788 (0.33701, 1.84196)	8/44	9/39
Brown et al., 2020	1.72222 (0.45075, 6.58021)	5/30	3/31
Duran et al., 2021	0.89161 (0.43441, 1.82999)	10/44	13/51
Litz et al., 2021	1.16291 (0.73592, 1.83764)	24/57	21/58
Lely et al., 2022	0.90000 (0.23781, 3.40606)	3/15	4/18
Vera et al., 2022	1.20000 (0.57259, 2.51491)	12/49	10/49
Ehlers et al., 2023	1.01905 (0.42106, 2.46627)	9/105	9/107
Litz et al., 2024	0.95187 (0.64120, 1.41308)	30/85	33/89
Dawson et al., 2018	1.00000 (0.02044, 48.91667)	0/32	0/32
Ahmadi et al., 2020	0.60000 (0.07447, 4.83420)	1/5	2/6
Bolton et al., 2014	1.35004 (0.83465, 2.18368)	32/114	21/101
Bryant et al., 2014	0.94615 (0.43086, 2.07773)	9/39	10/41
Crespo et al., 2010	0.84000 (0.33785, 2.08852)	6/25	8/28
Foa et al., 2013	1.03333 (0.22611, 4.72247)	3/30	3/31
Kangas et al., 2013	1.20000 (0.38860, 3.70565)	4/14	5/21
Sannibale et al., 2013	0.96286 (0.51324, 1.80638)	11/29	13/33
Shalev et al., 2016	0.90000 (0.56288, 1.43904)	16/40	28/63
Tay et al., 2023	1.31988 (0.36079, 4.82849)	5/161	4/170
Catani et al., 2009	1.06250 (0.02238, 50.43309)	0/15	0/16
Subgroup E ($I^2=0\%$, $P=0.71279$)	0.93377 (0.82076, 1.06234)	332/1750	386/1834
Sloan et al., 2018	0.55622 (0.35207, 0.87872)	21/100	37/98
Salloum et al., 2012	0.37374 (0.04084, 3.42048)	1/33	3/37
Subgroup G ($I^2=0\%$, $P=0.73031$)	0.54724 (0.34968, 0.85641)	22/133	40/135
Cohen et al., 2004	0.91304 (0.53719, 1.55186)	21/112	23/112
Nixon et al., 2012	1.06250 (0.43092, 2.61975)	6/16	6/17
Boelen et al., 2021	4.93333 (1.45870, 16.68460)	12/60	3/74
Cohen et al., 2005	2.00000 (1.11975, 3.57222)	22/41	11/41
Deblinger et al., 2011/1	0.92308 (0.61871, 1.37717)	24/52	26/52
Deblinger et al., 2011/2	1.04670 (0.68833, 1.59164)	25/54	23/52
Subgroup EK ($I^2=78.56\%$, $P=0.04971$)	1.30843 (0.78003, 2.19476)	110/335	92/348
Overall ($I^2=0\%$, $P=0.27805$)	0.94895 (0.85165, 1.05737)	464/2218	518/2317

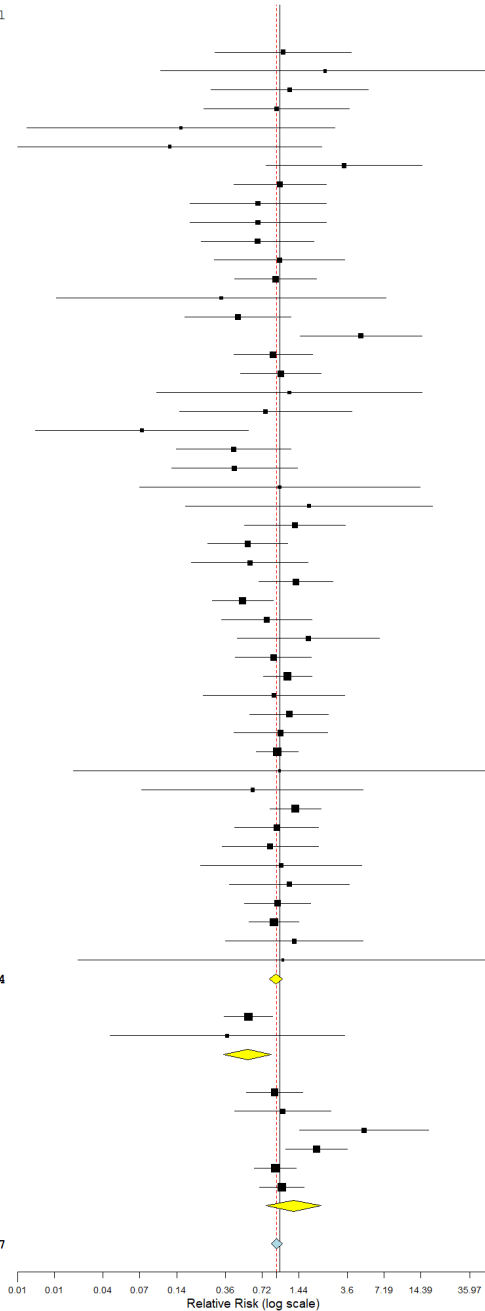


Abbildung 2: Abbrecherzahlen nach Behandlungssetting (E = Einzel, G = Gruppe, EK = Eltern-Kind), relative Risiken GFT versus TFT.

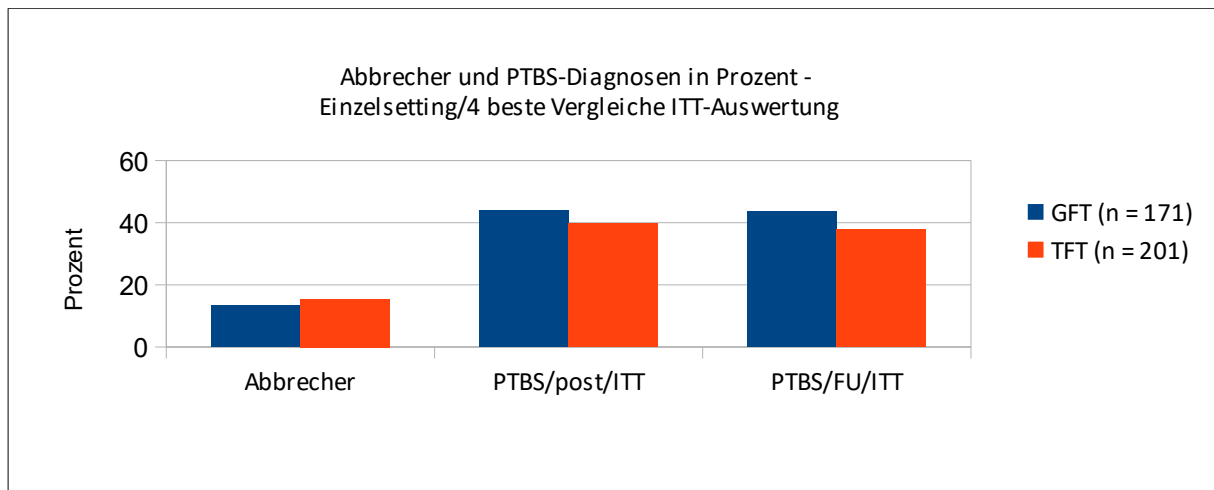


Abbildung 3a: Die nach Auswertung der 4 besten (QS 6) Vergleiche von GFT und TFT vorliegenden PTBS-Diagnosehäufigkeiten bei ITT-Auswertung zu den Zeitpunkten „post“ und „FU“.

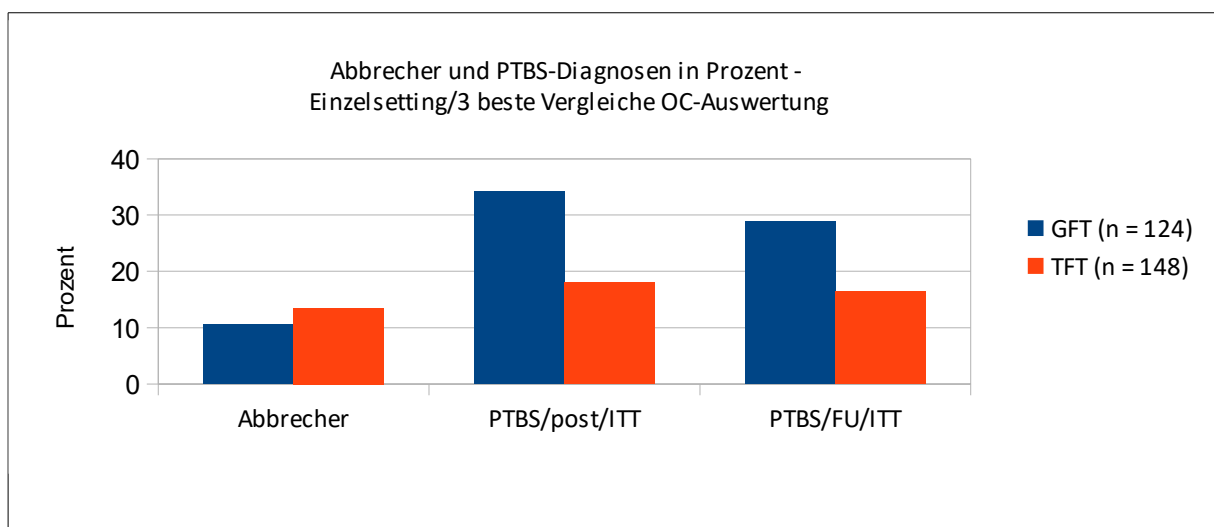


Abbildung 3b: Die nach Auswertung der 3 besten (QS 6) Vergleiche von GFT und TFT vorliegenden PTBS-Diagnosehäufigkeiten bei OC-Auswertung zu den Zeitpunkten „post“ und „FU“.

Diskussion und Schlussfolgerungen

Nach Auswertung von insgesamt 63 Behandlungsvergleichen scheinen die Belege für die Überlegenheit der TFT gegenüber der GFT überwältigend. Allgemein fallen alle als signifikant ($p < 0,01$) einzustufenden Unterschiede zugunsten der TFT aus. Die einzige Subgruppe, in der signifikant günstigere Werte für die GFT vorliegen, ist das Gruppensetting, in dem nach 2 Studien eine geringere Zahl an Behandlungsabbrüchen vorliegt. Im Gruppensetting wären allerdings allgemein ungünstigere Effekte zu erwarten (Bisson et al., 2013; Watts et al., 2013).

Auch wenn von einer Auswertung der besten Vergleiche verlässlichere und gültigere Ergebnisse zu erwarten wären, hat diese hier nur geringfügig veränderte Ergebnisse geliefert, mit weniger Signifikanzen, aber zum Teil etwas größeren Effektstärken als in der allgemeinen Auswertung und auch bei Tran und Gregor (2016). Ein starker Einfluss auf die Ergebnisse ist hinsichtlich der Researcher-Allegiance gegeben, womit sich ein dringliches Erfordernis zeigt, diese nicht nur bei der Interpretation von Ergebnissen zu berücksichtigen, sondern auch bei der Planung zukünftiger Studien, dies beispielsweise, indem die Forschungsteams von Vertreterinnen und Vertretern beider verglichener Behandlungen gebildet werden.

Dass hier stärker als in vergleichbaren Arbeiten zwischen den abhängigen Variablen differenziert wurde, hat als ein sehr gewichtiges Ergebnis erbracht, dass bei den traumafokussierten Behandlungsmethoden bei jenen Betroffenen, welche ihre Behandlung komplett absolvieren („observed cases“-Auswertung) eine fast doppelt so hohe Wahrscheinlichkeit besteht, dass sie zum Ende der Behandlung nicht mehr die Kriterien einer vollständigen PTBS aufweisen, als bei einer nicht traumafokussierten Behandlung. Mit einer einseitigen Einschränkung der Auswertung auf die „intent-to-treat“-Fälle oder eine Vermischung von abhängigen Variablen durch ein Mittelungsverfahren wäre dieses wichtige Ergebnis verloren gegangen.

Die Frage aber, ob nun die Größenordnung der festzustellenden Effektstärken als eher erheblich, mäßig oder doch nur geringfügig einzustufen ist, braucht damit nicht weiter diskutiert werden. Sie sind unterschiedlich. Doch sind die nach sehr strikten methodischen Kriterien gewonnenen Ergebnisse ausreichend, um die traumafokussierte Methoden in Übereinstimmung mit vorhandenen Empfehlungen (Schäfer et al., 2019) weiterhin befürworten zu können. Ein Dodo-Verdikt ist hier kaum mehr angebracht.

Interessenskonflikte: Die Arbeit und deren Inhalt stehen in keinerlei Zusammenhang oder Konflikt mit irgendwelchen wirtschaftlichen Interessen des Autors.

Literatur

Asarnow J, Ougrin D. From efficacy to pragmatic trials: does the dodo bird verdict apply? *Lancet Psychiatry*. 2017 Feb;4(2):84-85. doi: 10.1016/S2215-0366(16)30404-7. Epub 2016 Dec 1. PMID: 27914904.

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(5), 746–758. <https://doi.org/10.1016/j.cpr.2007.10.005>

Borkovec, T. D. & Nau, S. D. (1972). Credibility of analogue therapy rationales. *Journal of Behavior Therapy and Experimental Psychiatry*, 3, 257–260. [http://dx.doi.org/10.1016/0005-7916\(72\)90045-6](http://dx.doi.org/10.1016/0005-7916(72)90045-6).

Carroll, L. (1869). *Alice's Abenteuer im Wunderland*. Übersetzt von Antonie Zimmermann. London: Macmillan und Comp. https://de.wikisource.org/wiki/Alice_im_Wunderland.

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 (5), 553–563. <https://doi.org/10.1002/jts.20569>.

Klingler, O. J. (2023). *Seelische Verwundungen. Erscheinungsformen, Behandlungsmethoden und ein Selbsthilfeverfahren*. Norderstedt: Books on Demand.

Lewis, C., Roberts, N. P., Andrew, M., Starling, E. & Bisson, J. I. (2020). Psychological therapies for post-traumatic stress disorder in adults: systematic review and meta-analysis. *European Journal of Psychotraumatology*, 11 (1), 1729633. <https://doi.org/10.1080/20008198.2020.1729633>.

Luborsky, L., Rosenthal, R., Diguier, L., Andrusyna, T. P., Berman, J. S., Levitt, J. T., Seligman, D. A. & Krause, E. D. (2002). The Dodo bird verdict is alive and well – mostly. *Clinical psychology: science and practice*, 9(1), 2–12.

Luborsky, L., Singer, B. & Luborsky, L. (1976). Comparative studies of psychotherapies: Is it true that "everyone has won and all must have prizes"? In R. L. Spitzer & D. F. Klein (Hrsg.), *Evaluation of Psychological Therapies*, S. 3-22. Baltimore: John Hopkins University Press.

Munder, T., Flückiger, C., Gerger, H., Wampold, B. E., & Barth, J. (2012). Is the allegiance effect an epiphenomenon of true efficacy differences between treatments? a meta-analysis. *Journal of counseling psychology*, 59(4), 631–637. <https://doi.org/10.1037/a0029571>.

Rosenzweig, S. (1936). Some Implicit Common Factors in Diverse Methods of Psychotherapy. *American Journal of Orthopsychiatry*, 6 (3), 412-415.

http://clinica.ispa.pt/ficheiros/areas_utilizador/user11/6._some_implicit_common_factors_in_diverse.pdf.

Schäfer, I., Gast, U., Hofmann, A., Knaevelsrud, C., Lampe, A., Liebermann, P., Lotzin, A., Maercker, A., Rosner, R., Wöller, W. (2019). S3-Leitlinie Posttraumatische Belastungsstörung. Berlin: Springer. 155-001l_S3_Posttraumatische_Belastungsstoerung_2020-02_1.pdf (awmf.org).

Schnurr, P. P., Friedman, M. J., Engel, C. C., Foa, E. B., Shea, M. T., Chow, B. K., Resick, P. A., Thurston, V., Orsillo, S. M., Haug, R., Turner, C. & Bernardy, N. (2007). Cognitive behavioral therapy for posttraumatic stress disorder in women: a randomized controlled trial. *Journal of the American Medical Association*, 297 (8), 820–830. <https://doi.org/10.1001/jama.297.8.820>.

Schnurr, P. P., Friedman, M. J., Foy, D. W., Shea, M. T., Hsieh, F. Y., Lavori, P. W., Glynn, S. M., Wattenberg, M. & Bernardy, N. C. (2003). Randomized trial of trauma-focused group therapy for posttraumatic stress disorder: results from a department of veterans affairs cooperative

study. *Archives of General Psychiatry*, 60 (5), 481–489.
<https://doi.org/10.1001/archpsyc.60.5.481>.

Tran, U. S. & Gregor, B. (2016). The relative efficacy of bona fide psychotherapies for post-traumatic stress disorder: a meta-analytical evaluation of randomized controlled trials. *BMC Psychiatry*, 16, 266. <https://doi.org/10.1186/s12888-016-0979-2>.

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, 203–215.

Watts, B. V., Schnurr, P. P., Mayo, L., Young-Xu, Y, Weeks, W. B. & Friedmann, M. J. (2013). Meta-Analysis of the Efficacy of Treatments for Posttraumatic Stress Disorder. *Journal of Clinical Psychiatry*, 74 (6), e541-550.